Self-harm material consultation summary

Age-Restricted Material Codes

April 2026

Contents

Acknowledgments

eSafety acknowledges all First Nations people for their continuing care of everything Country encompasses – land, waters and community. We pay our respects to First Nations people and to Elders past and present.

eSafety acknowledges the meaningful and expert contributions of representatives from the following Australian agencies and organisations:

- National Suicide Prevention Office, Australian Government
- National Taskforce for Social Media, Body Image and Eating Disorders
- Orygen
- Butterfly Foundation
- ReachOut
- headspace
- Kids Helpline
- Beyond Blue

Thank you for sharing your expertise, perspectives and lived experience.

Disclaimer

This document represents a summary of views expressed during the consultation with each of these organisations. eSafety may not endorse in all circumstances all the information or its representation below. This summary **does not constitute legal advice, nor is it regulatory guidance from eSafety**. Service providers are responsible for ensuring the adequacy of their compliance actions, including consideration of their obligations under the Act, the Privacy Act, and any other applicable legislation or regulation.

Part 1: Introduction

The *Online Safety Act 2021* (Cth) (**the Act**) provides the eSafety Commissioner (**eSafety**) with legislative powers to help prevent end-users in Australia from being exposed to harmful online content and activity. This includes the registration and enforcement of Online Safety Codes and Standards, specifically the Unlawful Material Codes and Standards and the Age-Restricted Material Codes.

Overview

This document provides a summary of insights provided by consultation with Australian mental health services and working groups, with trusted knowledge and insights in the area of suicide, self-harm and eating disorders. This summary may provide online service providers additional information that may be considered when determining how to meet their obligations under the **Age-Restricted Material Codes** in relation to self-harm material. It also includes factors that service provider may consider in their determinations of material that may be considered as age-restricted self-harm material (**self-harm material**).

The intention of the **Age-Restricted Material Codes** is to establish appropriate safeguards to prevent children (under-18s) from accessing or being exposed to age-inappropriate content such as online pornography, high-impact violence material and self-harm material, while protecting the rights of adults to access it.

The intention of the **Unlawful Material Codes and Standards** is to prevent or restrict access to material that is unlawful or illegal and may pose harm to people of all ages. This may under some circumstances include suicide material, if the impact is such that it requires legal restriction.

When we refer to '**children**' in this consultation summary, we mean Australians under the age of 18. This summary may be read alongside the Online Safety Codes and Standards regulatory guidance, as well as eSafety's other corporate documents, including eSafety's compliance and enforcement policy.

Impact of self-harm material on children

The impact on children who encounter or engage with material that relates to self-harm on online services is complex and nuanced, and the potential for this material to be harmful can be highly contextual.

Research indicates that this type of content is relatively common on social media and other online services[1] and that many young people encounter this content.[2] There is also some indication that children are more likely than adults to be exposed to and create this material (such as self-harm and suicide content).[3]

Research highlights that not all forms of material which references self-harm is inherently harmful.[4] Some forms of content, such as content which conveys hope or provides help-seeking information, can be beneficial and valued sources of content for children. However, other forms of self-harm material may be more likely to be associated with harmful impacts or be harmful to some children or under some circumstances. In particular, content that may be of concern includes material promoting psychologically harmful ideas, maladaptive behaviours or behaviours harmful to health such as self-harm, suicide and disordered eating.[5] This includes content or communities that may promote, graphically depict, instruct, or otherwise encourage these behaviours.

Exposure to self-harm and suicide content may be distressing to encounter and has been associated with negative impacts to health, such as worsened mood and for some exposure was linked to further engagement in self-harm, including similar methods of self-harm to those depicted online.[6] Information shared to eSafety by Australian Coroners and by parents of children who have died by suicide, as well as those publicised in the media, also indicate patterns of engagement with self-harm material prior to death by suicide. Similarly, exposure to disordered eating content has been associated with body image concerns and other factors that may increase the risk of disordered eating.[7] Research and policy

---

[1] eSafety Commissioner (2025). The online experiences of children in Australia, eSafety website.
[2] Robinson, J., La Sala, L., Kenny, B., Cooper, C., Lamblin, M., Spittal, M., Gao, C., Kunin, M., Nicholas, A., Rezwan, A., Gifford, M., Pirkis, J. and John, A. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.
[3] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.
[4] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.
[5] American Psychological Association (2023). Health advisory on social media use in adolescence, APA website.
[6] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.
[7] Fardouly, J., Jarman, H.K., Bromberg, M., McLean, S.A., Wilksch, S., Fuller-Tyszkiewicz, M., Prichard, I., Mulgrew, K., Slater, A., Rossell, S.L., Griffiths, S., Maguire, S. and Marika Tiggemann (2024). Navigating Social Media, Body

recommendations highlight the need for online services to appropriately restrict or limit children encountering self-harm, suicide and eating disorder material that may be potentially harmful.[8]

Given the potential impact on the health and wellbeing of children, and the complexities in determining what content should be age-restricted self-harm material, eSafety undertook consultation through the form of written submissions and semi-structured discussions with expert stakeholder groups. eSafety sought feedback about how this type of material may be best described, and what other complementary steps online service providers can take to reduce harms (especially to children) who may be exposed to this material on their services. The consultation sought to build on existing terms and examples from similar international regulatory contexts in order to promote international regulatory coherence.

The importance of language

The way in which we communicate about eating disorders, self-harm and suicide has an important influence on how the community understands and responds to people experiencing these.

We acknowledge that language and preferences can be highly individual, and there is not always consensus on the how to best describe the range of experiences. eSafety also needed to work within predetermined frameworks in terms of how 'self-harm material' has been classified and conceptualised under the National Classification Scheme, and by the industry representatives who drafted the Age-Restricted Material Codes.

We acknowledge that the broad definition of 'self-harm material' has been used to describe what is often connected but distinctly different disorders, experiences or behaviours, and that these may be described and conceptualised in different ways.

Image, and Eating Disorders: Recommendations for Policymakers and Researchers to Drive Positive Change. doi: https://doi.org/10.2139/ssrn.4973962; Sala, A., Porcaro, L. and Gómez, E. (2024). Social Media Use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports*, [online] 14(14), pp.100404–100404. doi: https://doi.org/10.1016/j.chbr.2024.100404.

[8] Robinson, J., Thorn, P., McKay, S., Richards, H., Battersby-Coulter, R., Lamblin, M., Hemming, L. and Louise La Sala (2023). The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. *Frontiers in child and adolescent psychiatry*, 2. doi: https://doi.org/10.3389/frcha.2023.1274263

Sources of support

This consultation summary contains detailed information and descriptions about eating disorders, self-harm and suicide, and may be upsetting or distressing.

Where possible this material has been discussed sensitively and in only the level of detail required. This material has been developed with the needs of online services in mind.

If you experience, or if anyone you know is experiencing distress or requiring additional support, please ensure that you reach out to support services within your respective location. The following options outline some of the online and telephone support services available in Australia.

Lifeline **Phone:** 13 11 14 **Website:** lifeline.org.au	**Suicide Call Back Service** **Phone:** 1300 659 467 **Website:** suicidecallbackservice.org.au	**13YARN** **Phone:** 13 92 76 **Website:** 13yarn.org.au
Head to Health **Website:** medicarementalhealth.gov.au	**QLife** **Phone**: 1800 184 527 **Website**: qlife.org.au	**Beyond Blue** **Phone**: 1300 224 636 **Website**: beyondblue.org.au
MensLine Australia **Phone:** 1300 789 978 **Website:** mensline.org.au	**Kids Helpline** **Phone**: 1800 55 1800 **Website**: kidshelpline.com.au	**headspace** Phone: 1800 650 890 Website: headspace.org.au

Part 2: Defining self-harm material

Under the Head Terms to the **Age-Restricted Material Codes**, self-harm material is defined as:

> a subcategory of class 2 material defined for the purposes of [the Codes] as being comprised of **material** that is class 2 material because it **encourages, promotes or provides instruction** for:
>
> - **suicide**;
> - an act of deliberate **self-injury**; and/or
> - an **eating disorder** or **behaviour associated with an eating disorder**.[9]

The term material has the meaning defined in section 5 of the Act and includes written, video, audio and/or image-based material, whether it is real or synthetic (including AI-generated content). This material can be pre-recorded or shared live (such as via live streaming).

The Online Content Scheme

The Online Content Scheme in Part 9 of the Act is tied to the National Classification Scheme in terms of what may be considered 'class 1' and 'class 2' material and therefore falls within the jurisdiction of eSafety's regulatory remit.

Under the National Classification Scheme, the National Classification Board makes determinations about the rating of individual pieces of material as and when a request is made for a rating. In some circumstances, like for games and films, an automated tool may be used to determine a rating. The Scheme does not provide any tool or system for the classification of material which may appear on online platforms like social networks at scale.

In the process of developing the Age-Restricted Material Codes, industry representatives accepted the position proposed by eSafety in its Position Paper that suicide, self-harm and eating disorder material may be captured under the Australian Classification Guidelines as material within the 'Themes' classifiable element, as being material related to the themes of

[9] It is recognised that eating disorders are not traditionally accepted or conceptualised as 'self-harm' in other contexts.

suicide, death and serious illness.[10] This material is described in the Head Terms to the Age-Restricted Material Codes as 'self-harm material'.[11]

As none of these themes are explicitly defined under the National Classification Scheme, **we have consulted with expert stakeholder groups to understand how these may be best considered and described, to inform how service providers may action this material at scale**. This includes suggested definitions of key terms that have been devised based on feedback from our consultation with experts, and with consideration to existing terminology. To ensure regulatory cohesion, terminology has been aligned with comparable definitions internationally where similar content is regulated.

The following factors, highlighted through the consultation, have influenced the suggested definitions and examples outlined throughout this document:

- eSafety considered closely the terminology proposed by the United Kingdom online safety regulator Ofcom, which regulates in many cases the same or similar material as in Australia on the same global online services.[12] This terminology formed the basis of terms discussed in the consultations around definitional questions. The terms used by Ofcom were considered by consulted parties to appropriately capture 'self-harm material', with very limited variation needed for the Australian context. Accordingly, where appropriate, these terms and examples have been retained to maintain international regulatory coherence.
- Engagement undertaken with stakeholders to develop the 'National Suicide Prevention Strategy 2025-2035' indicated a preference for the use of the term 'self-harm' in comparison to 'self-injury'. However, there is a reference to self-injury in the definition of 'self-harm material' in the Age-Restricted Material Codes Head Terms. Given both self-harm and self-injury have distinct definitions under the Age-Restricted Material Codes the term self-injury has been retained to avoid confusion.
- It was also recognised through our consultation process that there are complexities and sensitivities with the terms 'deliberate' or 'intentionally'. The use of these terms was viewed as necessary to help online service providers to identify and address this material, however they are used with consideration that at the individual level this may not be reflective of a person's experience.

[10] eSafety Commissioner (2024). Development of Phase 2 Industry Codes under the Online Safety Act, eSafety website.
[11] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 2.1.
[12] Ofcom (2025). Protecting Children from Harms Online: Guidance on Content Harmful to Children, Ofcom website.

Key terms for self-harm material

Table 1 provides the descriptions for the key terms of suicide, self-harm and eating disorder material.

Table 1a: Guidance on key terms included in the definition of self-harm material under the Age-Restricted Material Codes

Term	Guidance
Suicide	An action that a person takes to deliberately end their own life.
Self-injury	An act in which a person intentionally harms themselves typically as a way of dealing with emotional distress, painful internal experiences, or overwhelming situations. An intentional act is an act that someone undertakes willingly or knowingly regardless of motive or intended outcome.
Eating disorder	Eating disorders are serious, complex and potentially life-threatening mental illnesses, which impact upon a person's physical health, mental health and holistic wellbeing. They are characterised by disturbances in behaviours, thoughts and feelings towards body weight and shape, and/or food and eating.
Behaviours associated with eating disorders	This includes behaviours that are generally associated with eating disorders across the various presentations. This non-exhaustive list includes the following behaviours: caloric restriction or fasting, binge eating, purging or other compensatory behaviours, avoidance or restriction of food, and excessive exercise.

Table 1b: Guidance on key terms in relation to types of material referenced in the definition of self-harm material under the Age-Restricted Material Codes

Term	Guidance
Material that promotes self-harm	Material which publicises, supports or recommends suicide, self-injury, or eating disorders (and associated behaviours). This also may include material that promotes the concealment or 'masking' or any form of self-harm. It is not required for this promotion to be intentional or explicit. This includes content which glamourises, glorifies, romanticises or normalises self-harm.
Material that encourages self-harm	Material which could incite or persuade others to contemplate or engage behaviours associated with suicide, self-injury, eating disorders, or behaviours associated with an eating disorder and/or make others more likely to attempt or consider this as a course of action. Encouragement does not have to be intentional or explicit. This can include material which glamourises, glorifies, romanticises, or normalises self-harm.

Material that provides instructions for self-harm	Describes or depicts a method or any actions that may be instructive in nature for suicide, self-injury eating disorders, or behaviours associated with an eating disorder, in sufficient detail that it can be emulated or replicated. Material does not need to deliberately or explicitly provide instructions for self-injury, suicide or eating disorders to be harmful. Instructions may be minimal and still able to be emulated. This could include describing or showing visually the materials or actions that might be undertaken that constitutes self-harm, such as setting out a 'plan' or steps, showing a diagram, providing 'tips', 'coaching' or 'guidance'.

Appropriate capture of material

As outlined in the Online Safety Codes and Standards regulatory guidance, material will only be self-harm material for the purposes of the Age-Restricted Material Codes where it is class 2 material.[13]

Some consultation participants raised concerns that the automated detection of self-harm material may inadvertently 'over-capture' and restrict access to essential health information and reduce the opportunities for the safe sharing of lived experience. eSafety recognises that online environments can provide beneficial access to community, mental health support services, and critical health information. **Service providers should undertake deliberate consideration to avoid disrupting access to beneficial mental health content, limiting engagement with reputable support services, or reducing the safe and appropriate sharing of lived experience.**

In addition, the consultation raised that the artistic representation of self-harm material can be harmful to end-users, particularly children. eSafety recognises that end-users, particularly those with lived experience, may wish to express their experience via different forms of artistic content, including music or lyrics, artwork, re-enactments or plays, stories, drawings or visual depictions, or fictional work that is shared or accessible online. **Service providers will need to carefully consider any artistic representations of self-harm material to ensure that the appropriate material is captured, particularly where it instructs, promotes, or encourages self-harm.**

The consultation also identified that self-harm material can **emerge and evolve rapidly**; it can be influenced by trend cycles and have unique terminology of relevance to this community. Additionally, those sharing or engaging with this content may present material in ways or that use specific language (such as code words, intentional misspelling or acronyms) that may not be detected by content moderation tools.

[13] eSafety Commissioner (2026). Online Safety Codes and Standards regulatory guidance, eSafety website.

Service providers should carefully consider and consistently refine their detection methods to ensure that the appropriate material is captured. Given these complexities and the rapid changes in the language and behaviours associated with sharing this material, working effectively with Australian support services is considered critical to effective responses to this material.

Part 3: Overarching considerations for services

Through consultation with services and individuals with expertise in the areas of suicide, self-injury and eating disorders, several themes emerged as relevant to making determinations about all forms of age-restricted self-harm material. These themes are summarised by eSafety in this section.

Individual impacts

Self-harm material can manifest online in a variety of ways and can impact children differently. Making decisions about this content can be complex.

Children may see the same content, but each child can be impacted by and respond to it in their own way. While all children are at risk when encountering self-harm material, some children are more likely to see this material and experience disproportionally harmful impacts. Often referred to as differential susceptibility, some children may be more influenced or impacted by material than others for reasons such as:

- individual differences in their lived experiences (such as exposure to stress or violence)
- experiences of circumstances of disadvantage
- experiences of trauma or abuse
- disability or neurodivergence
- temperament
- mental health
- age and development.

Given the design of many online services, children who are most susceptible or at greater risk of adverse impacts from engaging with self-harm material are also more likely to encounter this material[14]. Australian research found that younger people and those with a lived experience were more likely to be exposed to self-injury and suicide content even accidentally, in addition to being more likely to actively search for and create it.[15]

---

[14] Joint Select Committee on Social Media and Australian Society (2024). Social media: The good, the bad, and the ugly: Final report, Parliament of Australia website.

[15] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.

All actions taken by online services to prevent children's exposure and engagement with self-harm material should consider both responses that protect all children, and those required to protect children at greatest risk of harm.

Approaches that prioritise safety and provide users with greater control and agency over the content they are exposed to, including through settings and safety features, can help to ensure online experiences are more responsive to individual needs.

Sharing of lived experience and recovery material

'Recovery material' can relate to all types of self-harm material and may include, for example, content that focuses on how a person has recovered from an eating disorder. The consultation highlighted that recovery material needs to be considered carefully, as this can take many forms, and be both harmful and beneficial depending on the context, presentation and nature of the material.

It was identified that recovery material can be a source of hope, reduce stigma and encourage help-seeking behaviours. Simultaneously, material which is intended to be or shared under the guise of recovery-based material, can also be harmful. Relevant considerations regarding recovery material include:

- Some children will be more likely to be encouraged or instructed by some forms of recovery content regardless of the intention or nature of material.
- It may unintentionally be shared in ways that provide sufficient details to be emulated, glorifies, romanticises or promotes self-harm. This includes posts that while describing recovery also provide sufficient detail as to inadvertently instruct in self-harm methods or associated behaviours. Similarly, the inclusion of specific behaviours, quantitative metrics, or detailed comparisons (for example, numbers, frequency, duration, weight, or other measurable indicators) can contribute to copycat behaviours and harmful comparison, particularly in vulnerable children.
- Given the diverse range of individual experiences and vulnerabilities, material that may be helpful to one child, may be harmful to another. Subsequently, concepts such as how 'well-known' methods might be or promoting a 'less extreme method', were considered as challenging to apply to material at scale.
- Even when the primary material is not considered to be age-restricted, the harmful comments or reactions to the posted material may alter the potential for harm to be experienced.

Consultation participants highlighted a need to strike a balance between the safe and appropriate sharing of recovery content or lived experience, and the careful consideration

needed by service providers of the broader circumstances, as well as the potential for variability of children's experiences and potential for harm. As with other content, safe sharing should not provide detail that enables replication, include detailed instruction of the methods, glorify or romanticise, or promote in any other way. Messages of recovery were also considered to be most helpful when:

- emphasising the need to get appropriate help
- it was not sensationalised or aiming to generate engagement as the primary goal
- portraying realistic steps towards recovery (for example, not setting unrealistic expectations or unachievable goals).

The consultation further identified that there are a range of proactive education materials available that aim to help with the safe sharing of lived experiences and recovery information (such as the #chatsafe guidelines). Examples of these guides that can help people communicate safely about self-harm online are captured in Appendix A. While these are important resources, that should be made available to children and adult end-users, it is also recognised that the onus should not be placed solely on children to keep themselves safe online.[16]

Impact of design and digital features

Consistent with evidence, the consultation highlighted the role of design and digital features in enabling children to encounter and engage with self-harm material. In particular, the consultation identified that addressing recommender algorithms and generative artificial intelligence (AI) design features should be a key priority for service providers.

Recommender algorithms

The consultation identified that addressing recommender algorithms should be a key priority, given their role in the proliferation of harmful content.[17] Evidence suggests that harmful content may even be recommended without users actively searching for it or after minimal engagement.[18] Algorithms may also propel harmful combinations of

[16] Robinson, J., et al. (2023). The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. *Frontiers in child and adolescent psychiatry*, 2. doi: https://doi.org/10.3389/frcha.2023.1274263.

[17] Chhabra, J., Pilkington, V., Benakovic, R., Wilson, M., Sala, L.L. and Seidler, Z. (2025). Social media and youth mental health: a scoping review of platform and policy recommendations (Preprint). *Journal of Medical Internet Research*, 27. doi: https://doi.org/10.2196/72061.

[18] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.

content by recommending large volumes of content to the user, increasing the likelihood of harm to end-users. Due to the tailored nature of recommender algorithms, harmful content is also most likely to be presented to those who may be most impacted by exposure. It is recognised that this is further enabled when content moderation, user controls and reporting options are limited or absent.

The consultation also highlighted the role recommender algorithms play in the maintenance of mental health conditions, including exacerbating distress, reducing coping behaviours, and disrupting recovery. Given the tailored nature of recommender algorithms, the consultation reflected it can be difficult to have sufficient control over recommended content in a way that enables end-users to actively avoid content that may be harmful for them. An example was provided of how even when an individual with lived experience of an eating disorder (and who previously engaged with this material) attempted to proactively reduce their exposure to this content, they continue to be served material that was determinantal to their recovery.

Generative AI design features

The rise of generative AI services or features built into other services, such as AI chatbots, must also be monitored closely regarding their role in encouraging, promoting and instructing self-harm through interactions (such as sharing material and validating harmful behaviour) with end-users. Under the National Classification Scheme, material that is interactive may have a higher impact compared to content where exposure is shorter in duration, verbal and incidental, and not direct. Although the impacts of children using AI chatbots are still emerging, the consultation highlighted that these services have the propensity to contribute to the production and exposure of age-restricted self-harm material to children.

Part 4: Dealing with different types of self-harm material

A key consideration that emerged from the consultation is that there are several broader contextual factors which will increase or decrease the impact of self-harm material on end-users.

As the impact of material is a key factor that service providers must consider in aligning their compliance measures with the requirements of the Age-Restricted Material Codes, service providers may find it helpful to consider the following factors when assessing material.

Indicators that material may be lower-impact

Across all self-harm material types, consultation participants identified several factors which are typically indicative of lower impact on end-users in line with the National Classification Scheme. This includes where the material doesn't explicitly promote, provide instruction, include a level of detail that enables replication, or encourage suicide, self-injury, or eating disorder or behaviours associated with eating disorders in any way.

The consultation noted that while certain materials – such as those that explicitly promote self-harm – are more frequently associated with harmful impacts, some types of content may cause harm but not in all circumstances. The context can influence how likely material is to cause harm, including the characteristics of the specific nature of the material, and the broader circumstances upon which it is shared or encountered (such as the time of day).

A key consideration that emerged through the consultation was that material which is otherwise lower-impact may still have harmful impacts because of cumulative and combined influence from repeated exposure. For example, repeatedly serving material which promotes restrictive dieting practices and unrealistic appearance standards to an end-user via a discovery feed could be cumulatively harmful as it may contribute to the development or maintenance of disordered eating. The consultation highlighted that the cumulative harm of lower-impact content could be curbed by service providers having safeguards that also provide users with access to greater choice and control over the content they come across via feeds and recommender systems.

With that in mind, the consultation did consider some factors for service providers to consider, which may generally indicate that material may have a lower impact on end-users:

- **Material with limited detail:** Brief descriptions of self-harm where there are no details of the method, means, or without sufficient detail to provide instruction. The level of detail is sufficiently limited to prevent replication or emulation. Examples include:
    - A post where it is identified that someone died and that this was by suicide but does not otherwise describe the suicide.
    - An end-user sharing that they or someone they know have experienced self-injury but providing no detail of the methods or promoting it in any way.
    - A post that describes an individual journey of recovery from disordered eating. This may include reasons to recover, benefits of recovery, or encourages others to seek help.
- **Material that is academic, medical, statistical, clinical or educational**: This is material that is educational or informative in nature. This material may encompass a broad range of information types, including academic research or reports, health data, clinical definitions or descriptions, treatment information, position papers to direct legislative or policy interventions, prevention resources, or information to support self-management or recovery. Examples include:
    - Facts or statistics about suicide, such as the prevalence or associated risk factors, without detailed information regarding the method.
    - Information or research regarding the effectiveness of clinical interventions or treatment for self-injury.
    - Reputable health websites that describe risk factors or signs of an eating disorder or behaviour associated with an eating disorder.
    - Information or discussions around Voluntary Assisted Dying legislation.
    - Sharing beliefs or religious discussion about the afterlife.
- **Material that promotes safety, self-management strategies, help-seeking or recovery:** Broadly, any material that has a primary purpose of raising awareness, reducing stigma and increasing understanding, encouraging help-seeking, facilitating access to support services, promoting healthy coping behaviours, or promoting recovery in a safe and health promoting way, but does not otherwise promote, encourage or provide instruction. Examples include:
    - Guidance on how to communicate about suicide, or tips on sharing their lived experience, online in safe and non-stigmatising way.

- The sharing of information, resources, hand-outs or tip sheets related to self-injury on educational, medical or community health websites.
- Sharing about appropriate coping strategies and help-seeking avenues, including where someone may be able to receive medical or mental health support.

Intent of material

The consultation identified that the intent of the material does not preclude material from being harmful – for example, that recovery content may still be potentially harmful even if there are also help-seeking messages included with the material.

As discussed, recovery content requires careful assessment and consideration of contextual features – while it may in circumstances be beneficial for generating hope and reducing stigma, it also has the propensity to be harmful to children. Some of this material may be intended to support recovery, and others may be presented as such without this intent, the harmful content may also be in the response to the material such as within the comments.

Of paramount importance is that service providers should take the approach that children should not be inhibited from accessing credible and non-harmful resources that provides information about self-harm, treatment or support available, or that seeks to promote help-seeking behaviour.

Indicators that material may be higher-impact

The consultation participants identified that there are many factors which will increase the impact of self-harm material on end-users, making it more likely to fall within the scope of class 2 material. These factors are explored in this section, in relation to each category of self-harm material.

Material that might be considered age-restricted self-harm material will **often meet multiple definitions under age-restricted content**. For example, material that provides instruction may also promote and encourage self-harm. The examples provided in the following sections may also fall within more than a single category.

Considerations and examples for suicide material

Suicidal distress is a human response to overwhelming pain or suffering. Suicide or suicide attempts are not generally associated with an individual cause but arise from a dynamic interaction between factors specific to the person (individual and contextual factors), and their broader circumstances (social determinants). Some groups or communities are at a higher risk of suicide, due primarily to their increased experience of inequities and disparities in their experiences[19].

In addition to the potential for suicide material to be age-restricted, its promotion may also be considered as unlawful material under the Unlawful Material Codes and Standards or a criminal offence under some circumstances.[20] The Online Safety Codes and Standards regulatory guidance provides more detail on how material may be classified as age-restricted or unlawful, and service providers should refer to this when making those determinations.

Some material may be more readily identifiable as higher-impact, such as material that:

- provides detail of the method, allowing replication
- has pro-suicide messages, including glamorising, glorifying or romanticising suicide
- provides explicit instruction and/or graphic depiction.

Material that glamorises, glorifies or romanticises suicide

The consultation highlighted that material which glorifies, normalises, glamorises or romanticises suicide or other forms of self-harm requires careful consideration. In particular, the consultation highlighted that while material may not necessarily directly or explicitly promote or encourage suicide, the use of suggestive text or images or other accompanying content (such as hashtags, emojis, stickers, captions or filters) may all contribute to how this material manifests and, in turn, may be harmful.

This could also include material which portrays suicide in a positive manner (such as it being romantic) or describes it as a solution to problem (such as this was the only way to end their suffering).

[19] National Suicide Prevention Office. The National Suicide Prevention Strategy 2025-2035. Canberra: 2025.
[20] Schedule 1 *Criminal Code Amendment (Suicide Related Material Offences) Act 2005.*

The following table provides a non-exhaustive list of examples the types of materials that are likely to be considered as material which promotes, encourages or provides instruction for suicide.

Examples of suicide material
Material that depicts real-life or what appears to be real suicides or suicide attempts: • The sharing of an image or video where it is clear the person has died by suicide. This may include sharing of the method or other details that enable this to be determined. • Material that includes graphic descriptions or visual depictions of a suicide. This includes material that depicts a suicide attempt or where it is unclear if the individual has died or not. • Any material that depicts (including live streaming or pre-recorded video) a person preparing to take their life (such as discussing plans or showing their preparation related to the method) and/or the death of that individual. • An AI-generated video or image that uses the likeness of a person's image, and that acts out a suicide or suicide attempt.
Material that depicts or describes the methods or provides detailed instructions for methods of suicide: • Material containing descriptions or depictions of detailed steps to carry out a particular method for suicide. This may include a step-by-step guide. • Material that describes or depicts how to use, administer, prepare and/or procure any materials, items or substances associated with a method of suicide. • Material that describes the methods of how a person could attempt suicide or plans to take their own life. • Material that recommends locations of places where people have died by suicide or locations where this may be most likely to result in death. • Material that describes a method or methods of suicide as: being instant, more accessible, provides information relevant the threshold of pain (pain-free or painful), untraceable, easy to undertake, associated with high rates of death, or not having a lasting impact if it does not result in death. • Material that acts out or depicts the act of suicide in a way shows the methods or provides sufficient detail for replication. This includes content that is and is not life-like, such as animated content or content generated by AI. • AI chatbot providing details about suicide **methods, including methods** that are associated with the greatest likelihood of death or easier to enact with limited access to substances or materials.
Material that encourages or promotes suicide: • Material that dares or otherwise encourages users to engage in lethal self-harm or suicide methods. This may include suicide games, dares, online challenges or hoaxes. • Material – such as on discussions on forums - where end-users encourage each other to take their own lives. This may include 'suicide coaches', 'suicide mentors', or suicide pacts. • AI chatbot interactions that encourage users to engage in lethal self-harm or suicide methods, including suggesting methods that are least receptive to life-saving interventions, and/or provides communication which discourages appropriate help-seeking or treatment.

Considerations and examples for self-injury material

Self-injury refers to an act a person undertakes to harm themselves. People may engage in self-injury behaviour for many reasons. A common motive may be as a way of coping with distress or trying to manage intense emotions. Self-injury is a risk factor for suicide, and may result in death even when the person had no intent to end their own life.

There are a range of behaviours or actions that may be considered as self-injury. eSafety recommends service providers work closely with mental health services and prevention experts to determine the behaviours that may be categorised as self-injury.

Risk-taking behaviours

The consultation highlighted that young people's interpretation of self-injury may be inclusive of a range of behaviours that may be harmful to health, that are broader than those traditionally included in this description. This may include risk-taking behaviours such as binge drinking, dangerous driving or drug use.

Although it was considered beneficial that young people are provided sufficient control over their online experiences to minimise exposure to all these forms of content, eSafety considers the traditional conceptualisation is most aligned with the regulatory requirement to prevent exposure to class 2 material.

The consultation considered that there are some examples of material, specific to self-injury, that may be unlikely to be considered as age-restricted self-harm material. Consistent with the considerations in other jurisdictions, this includes material that show:

- healed scars that are visible incidentally (without any active wounds)
- how to cover healed scaring but not active concealment of harm
- safe and health-promoting coping strategies or management of negative emotions (such as recognising triggers or engaging in preventative coping techniques).

The following table provides a non-exhaustive list of examples the types of materials that are likely to be considered as material which promotes, encourages or provides instruction for self-injury.

Examples of self-injury material

Material that shows real, or what appears to be real, self-injury:

- Any material that depicts (such as live streaming or pre-recorded video) or that shows a person preparing to harm themselves (such as discussing plans or showing their preparation related to the method) and/or the act of self-injury.
- Material that includes graphic descriptions or visual depictions of an act of self-injury.
- Material that shows images of self-injury or active self-injury wounds presented without appropriate context (for example, in a clinical or medical context). This may also include healed self-harm injuries if shared in a way the promotes, encourages or instructs in self-injury.
- Material that depicts the injury or sustained impact of self-injury. For example, photos or videos that compare body parts before and after.
- AI-generated videos or images that depict or show an act of self-injury.

Material that describes methods or contains detailed instructions or actions that can be replicated:

- Material that describes or depicts a method of self-injury with sufficient detail to be emulated. This could include describing how they have, how they intend to, or how someone could harm themselves. This includes content that is and is not life-like, such as animated content or content generated by AI.
- Material that describes or demonstrates the specific tools, substances and/or items used for self-injury. This may include how to use them or where to obtain them.
- Material – such as AI chatbot interactions – that provide information about ways in which objects may cause harm when used in a particular way, describes as easy to do, easy to conceal, or quick to heal.

Material that encourages or promotes self-injury:

- Material or comments which encourage competition, may incite increased acts of self-injury, or encourage the escalation of self-injury. This includes direct encouragement to engage in self-injury, suggesting that the method or impact is insufficiently severe, or providing suggestions of other or more harmful methods.
- Material that depicts accessories that are commonly associated with self-injury which is accompanied by additional content (such as text or sounds) or another content modifier (such as a hashtag).
- Material that romanticises, glamourises or normalise self-injury, such as describing it as poetic or as a healthy response. This may include sharing of lived experience, artwork, fanfiction and AI-generated content.
- Material that dares, challenges, or otherwise encourage users to engage in behaviour which may amount to self-injury methods. For example, an online challenge that encourage the ingestion of a harmful or potentially harmful substance.
- AI chatbot interactions that encourage users to engage in self-injury, recommends or validates these actions, and/or which discourages appropriate help-seeking or treatment.

Considerations and examples for eating disorder material

Eating disorders are complex and serious mental health conditions that can have potentially life-threatening impacts. There are several types of eating disorders (for example Anorexia Nervosa, Bulimia Nervosa, Binge Eating Disorder), each associated with specific but also overlapping characteristics and behaviours. Eating disorder presentations can also be expressed in a range of ways across different communities or groups of people (such as differences across genders). The consultation **recommended that service providers work closely with recognised experts** to understand how the differences across these eating disorders and impacts on specific groups may influence the types of content that could be potentially harmful to children.

Childhood and adolescence are vulnerable developmental periods for the onset of body image concerns and eating disorders, with body dissatisfaction being a strong predictor of eating disorders[21]. There is a broad range of content that may be potentially harmful for children at risk of body image concerns or eating disorders. This can range from more explicit or extreme content such as pro-eating disorder material to more indirect or cumulative influences that normalise unrealistic standards and narrow beauty ideals, stigmatise people based on their weight or appearance, or focus excessively on dieting, weight-loss and exercise. Determining age-restricted eating disorder material can also be challenging as harmful content may be embedded in content that is or is presented as recovery-focused or shared under the guise of 'healthy' lifestyles or 'fitness' content.

As discussed, the impact of this material may be further propelled by algorithms that both extend time online and enable the targeting of material to end-users, particularly to those most at risk of harm. Experts identified that this could exacerbate harmful impacts by repeatedly exposing children to age-restricted material, as well as exposing them to harmful combinations of material (such as appearance-focused content, pro-eating disorder content, and weight loss related advertising).

The impact and harmfulness of eating disorder material may be influenced by a range a factors, including the tone, the underlying message or framing, or additional supporting content (such as captions, pro-eating disorder quotes, hashtags or audio) and the associated imagery. The consultation raised that even material which at first seems harmless or recovery-focused may become harmful when all these factors and the context are considered.

[21] Fardouly, J., et al. (2024). Navigating Social Media, Body Image, and Eating Disorders: Recommendations for Policymakers and Researchers to Drive Positive Change. [online] doi: https://doi.org/10.2139/ssrn.4973962.

The consultation also highlighted these impacts can be further amplified by beauty filters that normalise and promote the production of visual content that reflects unattainable appearance standards. Adolescents often develop 'parasocial relationships' with influencers, in which they tend to view them as trusted role model or a friend, even in the absence of any reciprocal interaction.[22] Influencers can be particularly persuasive to adolescents, including shaping health beliefs and behaviours. This can leave adolescents particularly susceptible to influencer-driven advertising.[23]

Examples of eating disorder material that may be harmful

Some examples of content or indicators of material that may be more likely to be associated with harmful impacts provided in the consultation included:

- Material that shows before and after image (or other forms of comparison images) that focus on showing bone structures of body parts associated with a clinically unhealthy BMI (such as ribs, collarbones and thigh gaps).
- Material that is descriptive and includes measurable details (such as weight or body measurements) which promotes comparison, material that provides a level of detail that enables emulation (such as how much or little is eaten or methods of purging).
- Material that encourages or invites comparison and/or affirming feedback of extremely low body weight or body parts, for example, 'body checking' content.
- Material that is shared to incentivise, encourage or to motivate others to engage in eating disorders or associated behaviours, or aspires others to replicate the impacts of eating disorders or associated behaviours. This includes 'thinspiration'.
- Harmful material may also be in the comments or reactions, as a response to material shared by a person or others.

Conversely, there were risks identified with the potential over-capture of health information or safe recovery content where blanket restrictions were in place (such as restricting all mentions of eating disorders). Service providers will need to identify moderation approaches that ensure the safe sharing of images of people with low body weight are not restricted.

[22] American Psychological Association (2025). APA Recommendations for health teen video viewing: A summary of the science with action steps regarding video viewing and adolescent well-being. APA Recommendations for Healthy Teen Video Viewing

[23] American Psychological Association (2025). APA Recommendations for health teen video viewing: A summary of the science with action steps regarding video viewing and adolescent well-being. APA Recommendations for Healthy Teen Video Viewing

For example:

- Material that shows people of low body weight, where they are not glamorising, instructing, encouraging, normalising or promoting eating disorders in anyway, would not be considered to be included as age-restricted self-harm material.
- Relatedly, services should not restrict material which shows diverse eating practices that do not otherwise promote or encourage disordered eating, such as relating to religious practices or occasional overeating in isolated circumstances (such as for a celebration).

The consultation also highlighted that although the examples of material focus on emaciation or the 'extreme thinness end', people of any body shape and size can experience an eating disorder. That material may still promote eating disorders even in the absence extremely low body weight or clinically unhealthy BMI. Additionally, there were concerns identified with online material that promoted binge eating or high-caloric consumption, both in the context of an eating disorder and in dangerous online challenges that amount to behaviours associated with an eating disorder.

The following table outlines a non-exhaustive list of examples of material that may be considered to instruct, encourage or promote eating disorders or behaviours associated with eating disorders.

Examples of eating disorder material
Material that describes with sufficient detail to be instructive or includes instructions for methods related to eating disorders or behaviours associated with eating disorders: • Material that provides instructions or depicts the methods in a way that it may be replicated or emulated. This can be in across a range of formats, such as in posts, blogs or story content. It may include methods of purging, compensatory behaviours, food or calorie restriction, or excessive energy expenditure. This includes content that is and is not life-like, such as animated content or content generated by AI. • Material which details how certain types of substances, pharmaceutical products, or weight loss related products may be procured and the way they may be taken or used. • Material – such as interactive material with an AI chatbot – that provides information about how eating disorder or associated behaviours might be concealed or hidden from others. • Material that describes or depicts behaviours that amount to eating disorders or associated behaviours. This may be inadvertent or shared in a recovery context.
Material which promotes or encourages eating disorders or behaviours associated with eating disorders: • Material that provides specific detail relating to an eating disorder. Such as calorie counts, BMI or body weights, or other forms of measurement. • Material that promotes the use of any substance, medication or other product in the context of an eating disorder or associated behaviour. • Material that indicates or promotes symptoms or impacts as desirable or a measure of success (such as the ability to restrict calories to an extremely low level).

- Material that incentivises eating disorder behaviours by encouraging disgust or weight stigma and discrimination towards bodies which are not underweight.
- Material that uses negative feedback, shaming or criticism as an incentive to engage in eating disorder behaviours or to promote an escalation or engagement in more extreme or severe methods (such as 'meanspo'). This may be requested by an end-user for the purpose of motivation, or it may be provided by others who criticise a person's behaviours or their methods as not 'extreme enough'.

Material which idealises, glamorises, or otherwise presents these behaviours or the associated impacts as aspirational or desirable in nature:

- Material which is shared as motivation to encourage or glamorises eating disorders. This includes material that glamorises extreme thinness or otherwise presents extremely low or clinically unhealthy body weight or BMI as aspirational (such as 'thinspiration' material).
- Material that romanticises extreme thinness and emphasises specific body parts such as protruding collarbones, spines, scapulas, hipbones, rib bones, or flat or concaved stomachs, or 'thigh gaps'. This material may be presented in the form of a 'body check'.
- Other forms of material which glamorises or romanticise dangerously low body weight or clinically unhealthy BMI. This may include artistic content that is audio (such as lyrics), text (such as poetry or motivational quotes) or image-based (such as artwork or AI-generated material). Particularly when accompanied by content modifiers or interactive features which actively connect it to known pro-eating disorder hashtags.

Material which promotes or depicts extreme or excessive weight loss, dieting, or fitness content:

- Material which provides instructions for weight loss, dieting or exercise which is extreme and/or excessive and/or dangerous. This may include drinking water to curb hunger or persistent calorie restriction.
- Material that emphasises unsafe, unhealthy or extreme weight loss transformation. This material may be in the form of before and after weight loss transformation. This includes weight loss where the after-picture shows either bone structures or clinically unhealthy BMI and/or mentions a rate of weight loss that would be unhealthy or dangerous. These may be accompanied by specific weights, BMI or measurements.
- Material which promotes or encourages substances, medications or other products to promote excessive weight loss or dieting.

Material which encourages interactive behaviours associated with eating disorders (including in pro-eating disorder or community forums):

- Material that is shared with other end-users – such as in community groups, forums or group messages – that promote eating disorders or associated behaviours, such as acting as a 'coach', 'mentor' or 'buddy', or encourage engagement in behaviour as a group.
- Material that encourages others to share images or videos of age-restricted self-harm material (such as images of purging).
- Sharing or requesting 'tips', 'hacks' or 'motivation' for behaviours associated with eating disorders, with users listing goal weights, current weights and/or lowest weights.

This may also include challenges related to eating disorders or associated behaviours, such as:

- Challenges that encourage extremely low-calorie consumption, losing weight in extreme ways or amounts, or in relation to body shape or size (such body comparison challenges). This can also include, though less frequently, challenges that encourage excessive caloric or food consumption.

Part 5: Crisis information and support services

The provision of crisis and support services information has been identified in the research and in the consultation as an important safety feature for encouraging help-seeking and promoting safer online experiences.[24] This feature was found to be particularly helpful for children and people with a lived experience.[25]

In addition to being beneficial, this may be required through compliance measures of particular Age-Restricted Material Codes. When providing this information or support, the consultation highlighted that in general this should ideally:

- include direct links to Australian services, and where possible (there is sufficient information to determine the nature of the concern) provide at least one support services relevant to the specific concern identified
- ensure that options are accessible – this includes options that are free, open to all ages and those that are specific to the age of the user and provide 24/7 support services
- provide a range of modalities where possible, including text and professionally moderated online forums in addition to call-based services, and that these various options are identified
- provide services that may be more beneficial or inclusive for community groups, so that people can reach out for support in a way that suits their specific needs. This includes designated services for youth, First Nations people, the LGBTIQ+ community, people with disability, people who are culturally and linguistically diverse, and those who are neurodivergent.

Service providers must use appropriate judgment to balance the need to provide a range of services to enable choice, while keeping this list simplified and focused to only key services. Where the online service has information about a user, such as their age, that might be helpful in determining most appropriate services this should be used.

---

[24] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.

[25] Robinson, J., et al. (2025). How do Australian social media users experience self-harm and suicide-related content? A National cross-sectional survey comparing young people and adults. *BMC Public Health*, 26(1). doi: https://doi.org/10.1186/s12889-025-25646-0.

Considerations for providing crisis support and information

In addition to the types of information or services that should be provided, the consultation highlighted that other additional characteristics that are important in encouraging help-seeking. The consultation group suggested the following:

- Use of language and terminology that is **accessible** and **appropriate** to the comprehension and developmental needs of those accessing the service.
- Services should use visual cues, plain language and inclusive messaging to increase accessibility.
- Emphasis on the importance of getting support, highlighting that there are resources available and that a child should act on this without delay.
- Use of language that is **neutral,** broadly encourages help-seeking, and appropriate for children when discussing themes that may be potentially harmful.
- That messaging uses a **youth-friendly tone** and is personalised to a child or their experience.
- That there are **clear actions** for children to follow to get help or support. This may also include suggestions to access broader health support options such as suggesting they speak with their general practitioner (GP).

eSafety recommends service providers work with mental health organisations or relevant experts to ensure that these factors are considered and implemented in an effective way.

Promoting trustworthy sources of information and support

It was evident from the consultation that online services are an important source of health information and support for young people – however, there are concerns about the quality of some health and mental health information available and promoted to children.

Indicators of trustworthy sources

The consultation highlighted some indicators of trustworthy sources of information or material:

- Original source material is created by a licenced health professional or recognised mental health organisation, research body, institute, peak association, university, advisory or academic group, or other entity with recognised accreditation or expertise.
- The material provides appropriate evidence-informed sources, based on peer-reviewed research where available, with a focus on providing factual information over generating engagement.

During the consultation, entities with recognised experience in the areas of self-injury, suicide and eating disorders expressed a willingness to support online services to identify a list of relevant sources. The consultation highlighted that this includes ensuring that health information, regardless of whether children encounter this following active engagement (such as searching for information), or more passively or unintentionally encountered (such as through recommended content or accounts), is accurate, trustworthy and evidence-informed.

In alignment with evidence, the consultation highlighted that young people also express that they desire ways to more easily identify credible sources of mental health information.[26] Further to this, the consultation highlighted that service providers should consider how to verify and label trustworthy and authoritative health content. This included verified health content being promoted algorithmically over non-verified sources.

Identifying appropriate Australian Services

A list of safety and support services relevant to self-harm material, identified through the consultation, is included in Appendix A. This list is not exhaustive and is a point-in-time of the production of this summary. Most appropriate services may change over time and depend on the needs of the service. This list therefore does not represent a recommendation on support services to meet compliance. Online service providers are responsible for ensuring that the services they provide are appropriate to the purpose, they provide services that are contemporary, and that meet the requirements in the codes, particularly as these may evolve over time.

[26] ReachOut Australia (2024). How young people use social media for mental health information and support, ReachOut Australia website.

In addition to crisis information, service providers should consider the benefit of proactively linking out to resources that help young people create material or contribute in ways that are less likely to generate age-restricted self-harm materials. For example, people who are posting regularly in these forums or about these topics might be directed to or promoted resources that help them understand beneficial and safe ways to communicate about complex and sensitive topics online. Some examples of these are provided in Appendix A.

Considerations for working more effectively with Australian support services

Consistent with the research, the consultation identified that evidence around how online services may best manage or respond to self-harm content is still emerging and highly contextual.[27]

Working with reputable Australian mental health, community support services and prevention experts will help to ensure that online services are providing contemporary, evidence-informed, and contextually relevant information and resources to children. Consulted prevention experts and mental health services expressed a willingness to work with online services to create safer online experiences for children. It was also reflected that some online services already work closely with Australian mental health services and prevention experts.

Opportunities to strengthen this engagement – highlighted in the consultation – and to ensure that it is meaningful and effective include:

- **Transparency**: This includes providing timely feedback on how information shared during consultation will be used, why recommendations may not be implemented, and the impact of these recommendations. Service providers could do this by voluntarily publishing annual transparency reports that are publicly available.
- **Co-design:** Children and other diverse groups should be involved in the co-design of products or features. Their needs and expressed views should be appropriately considered and have demonstratable influence.
- **Early and ongoing engagement**: Consultation with stakeholders should be proactive, including them in the design, development and testing phases of products and features.

[27] La Sala, L., Sabo, A., Michail, M., Thorn, P., Lamblin, M., Brown, V. and Robinson, J. (2024). Online safety when considering self-harm and suicide-related content: Qualitative focus group study with young people, policymakers, and social media industry professionals (Preprint). *Journal of Medical Internet Research*, 27. doi: https://doi.org/10.2196/66321.

Furthermore, that service providers continue to engage with stakeholders to ensure effective implementation of recommended actions.

- **Responsible data sharing**: That consulted parties, particularly those with research capabilities, can access relevant data to understand and evaluate impact, as well as identify opportunities for additional safeguards.
- **Diversity in perspectives:** In addition to consultation with reputable and national mental health organisations, peak associations and research institutes, that service providers also ensure they include a diverse range of perspectives in their consultation. This includes children and young people, First Nations people, culturally and linguistically diverse communities, LGBTIQ+ communities, people with a disability, recovery advocates, those with a lived or living experience mental ill-health, and those with clinical expertise.

Promoting and maintaining effective partnerships with mental health services and prevention experts should be a priority for online services.

Part 6: Safety tools and continuous improvement

The consultation highlighted that there are a range of safety tools, safeguarding practices, and design approaches, that can help to address and reduce children's exposure to and engagement with harmful content.

Many of these recommendations align with compliance measures in the codes, such as the requirement to implemented safety tools, whereas others although outside the requirements in the Online Safety Codes may be considered by services in their effort to continuously improve the safety of children accessing their services.

Algorithms identified as priority for improvement

The need to address recommender systems and algorithms, as a key mechanism that proliferates and amplifies children's exposure to self-harm material, was identified as a priority area for continuous improvement in the consultation. Consistent with research, it was considered that service providers should not only limit children's exposure to this content but also ensure that online services do not recommend (such as in suggested content) or drive users to this content.[28]

The consultation also highlighted existing research that indicates harmful content is not only propelled by algorithms but enabled by limitations in content moderation systems and ad management systems.[29] Recommendations highlighted the role of both AI moderation and human review in effectively preventing children from encountering self-harm material.

Furthermore, in addition to reducing harmful content, it was identified that algorithms can be shaped to promote material that may be beneficial or health promoting. Examples included prioritising and promoting trustworthy health content, stories of hope, and promoting greater diversity in appearances online.

---

[28] Robinson, J., et al. (2023). The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. *Frontiers in child and adolescent psychiatry*, 2. doi: https://doi.org/10.3389/frcha.2023.1274263.

[29] Reset Tech Australia (2024). Not just algorithms: assuring user safety online with systemic regulatory frameworks, Australian Policy Online website.

Consultation participants identified that the implementation of safety tools or settings to limit access or exposure to self-harm material could include:

- Filtering, limiting or blocking self-harm material
    - This could include providing users with a 'shield function', which enables them to opt-out of encountering self-harm material or other forms of content that may be harmful to or undesired by the person.[30]
    - For children, access to age-restricted self-harm material should be blocked by default, as required under the codes. A shield function may provide additional benefits by enabling children to limit exposure to self-harm content that is not age restricted and other content which may cause harm to an individual (such as dieting content, fitness content).
- Removing self-harm material from targeted advertising
    - For example, service providers could give the users the ability to opt-out of targeted advertisement (particularly for appearance related advertising) and that for children as a default this is should not be enabled.
- Greater control over recommender systems and algorithms
    - This should include allowing end-users to reset their recommender algorithms easily and on demand. Services should also provide clear and accessible information about the effect and duration of the reset to users.
    - This includes providing setting options – which should be enabled by default – for recommender systems that promote long-term positive value rather than prioritising engagement.
- Limiting cumulative exposure via endless content streams
    - This should include features such as auto-play and infinite scroll turned off or disabled by default. Services should also provide clear and accessible information about the risks associated with using these features and accompany any use with appropriate safeguards (such as automatically turning off after a period of use).
- More comprehensive and effective reporting tools
    - Consistent with research, to effectively reduce exposure to harmful content, service providers should ensure that reporting methodologies do not limit what

[30] The term 'shield function' was coined and recommended by the National Taskforce for Social Media, Body Image, and Eating Disorders. Refer to Fardouly, J., et al. (2024). Navigating Social Media, Body Image, and Eating Disorders: Recommendations for Policymakers and Researchers to Drive Positive Change. [online] doi: https://doi.org/10.2139/ssrn.4973962.

end-users are able to report – for example, if a service provides reportable categories, these categories should encompass a broader range of content.[31]

- That content which is reported as harmful to the end-user should be appropriately actioned as required under the Age-Restricted Material Codes. This could include providing end-users with the option to see less of that type of content – even if the material is not removed.

Use of sensitive content warnings and click-throughs

The consultation highlighted the evidence on the effectiveness of tools relating to sensitive content warnings, image blurring and click-through warnings is inconclusive and/or mixed.

Despite the evidence on the effectiveness being unclear, the **consultation highlighted that these are often tools that are requested and valued by young people in shaping their online experiences.**

The consultation also raised concerns that these tools, if not applied effectively, particularly to the safe sharing of lived experience, may increase stigma or leave end-users feeling invalidated. Online service providers should work closely with Australian health services to consider the most effective ways to promote, offer and implement these tools.

Additionally, consultation participants highlighted the importance of the ensuring the effectiveness of these safety tools. The consultation highlighted some considerations for these tools, which may include the following:

- Service providers should ensure that safety tools are **easy to use, accessible and developmentally appropriate**. Service providers should keep end-users informed about these safety tools. For example:
    - Users should regularly be made aware of the safety tools available to them and how they can be activated, including when these safety tools are updated.
- The highest safety settings should be enabled by default. Where only the highest safety setting is developmentally appropriate, safety settings may not be adjustable – to less safe alternatives – for children.

[31] Robinson, J., et al. (2023). The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. *Frontiers in child and adolescent psychiatry*, 2. doi: https://doi.org/10.3389/frcha.2023.1274263.

- When switching off safety settings, end-users should be provided with relevant warnings relating to health and safety.
    - That information about how to use safety features and customise safety settings is included in crisis prevention information.
- To ensure that these tools are effective, service providers should regularly evaluate these tools. This includes the extent to which end-users are aware of the tools, have access to these tools, and that provide a meaningful difference in their online experience.
- Service providers should ensure that they **meaningfully engage with mental health professionals and experts** throughout the product development lifecycle so that safety tools are informed by evidence and are effective. This includes:
    - Appropriate consultation during the development and testing of product requirements for successful implementation.
    - Ongoing transparency with data that will best ensure online services are providing safe and health-promoting services.

Appendix A: Example list of support services

Crisis support services[32]			
Kids Helpline	5 to 25 years	Free, confidential 24/7 online and phone counselling service.	Call 1800 55 1800 Online chat
Lifeline	All ages	Free, confidential 24/7 online and phone counselling service.	Call 13 11 14 Text 0477 13 11 14 Online chat
13YARN	All ages	Free, confidential 24/7 phone counselling service for Aboriginal and Torres Strait Islander peoples.	Call 13 92 76
Suicide Call Back Service	15+ years	Free, confidential 24/7 online and phone counselling service to people who are at risk of or affected by suicide.	Call 1300 659 467 Online or video chat

Counselling and specialist services (various ages)			
Butterfly Foundation	All ages	Support for eating disorders and body image issues. Services from 8am to midnight, 7 days a week.	Call 1800 334 673 Online chat
ReachOut	16 to 25 years	PeerChat is a one-on-one, text-based service with a peer worker. It is available Monday to Thursday 3pm to 8pm, Friday 11:30am to 4:30pm AEST.	Online chat
Beyond Blue	All ages	Free, confidential 24/7 counselling services and online forums.	Call 1300 22 4636
headspace	12 to 25 years	Free, confidential and anonymous mental health support for young people and their family. Available 3pm to 10pm every day.	Call 1800 650 890 Online chat
Qlife	All ages	QLife provides anonymous and free LGBTIQ+ peer support and referral for people in Australia. Available every day from 3pm to 9pm.	Call 1800 184 527 Online chat

[32] A person who is at immediate risk in Australia harm should always been directed to Triple Zero (000) for emergency services. Although this may be presented alongside other crisis services, this should be visually prioritised.

Standby Support After Suicide		StandBy is dedicated to assisting people and communities bereaved or impacted by suicide. The StandBy program supports anyone who has been bereaved or impacted by suicide. Provides a range of free services, including Peer Support and Suicide Bereavement Counselling.	Call 1300 727 247 Website

Counselling and specialist services (adults)			
Sane	18+ years	Provides a range of free, complex mental health support, including phone or online chat (available Monday to Friday, 10am to 8pm), the RecoveryClub program with groups and 1:1 support, 24/7 community forums, events, and information and resources.	Explore all support
MensLine	15+ years	MensLine Australia offers free professional 24/7 telephone counselling support for men with concerns about mental health, relationships, anger management, family violence (using and experiencing), stress, and suicidal thoughts.	Call 1300 78 99 78 Chat or video call online
Medicare Mental Health services	Service navigation	Medicare Mental Health Centres and Kids Hubs can connect you with a range of mental health professionals for support, whatever your situation.	Medicare Mental Health services
Open Arms	18+ years	Offers face-to-face, telephone and online counselling to anyone who has served at least one day of continuous fulltime service in the Navy, Army or Air Force, and their immediate families. Hotline is available 24/7.	Call 1800 011 046 More information
Defence Member and Family Helpline	Australian Defence Force members or family	The Defence Member and Family Helpline is the first point of call for Defence families seeking support, information or connection with their community.	Call 1800 624 608 More information about services

Self-management tools and moderated forums			
InsideOut Institute Eating Disorders	16+ years	The eClinic provides free digital therapies for people experiencing eating or body image concerns. All programs are available immediately, require no referral, and can be completed independently at your own pace or with support from your health professional. The platform also includes a dedicated program for carers and families, as well as a specialised hub for health professionals.	InsideOut eClinic
Butterfly Foundation	16+ years	The Butterfly Collective is an online lived experience network involving people across Australia who either have a lived experience of an eating disorder or body image concerns or are a carer, family member, or friend of someone with a personal experience.	Lived Experience Network – Butterfly Foundation
headspace	12 to 25 years	Offers moderated group chats on topics like relationships, anxiety, and self-care for young people and their families.	headspace online communities
Kids Helpline	12 to 25 years	My Circle is the free, private, safe and confidential social platform for 12- to 25-year-olds across Australia.	My Circle

Resources for communicating about self-harm online		
Orygen	A young person's guide to communicating safely online about self-harm and suicide.	#chatsafe guidelines
InsideOut Institute for Eating Disorders	A guide to communicating safely online about eating disorders and sharing stories of eating disorders with others.	Creators Guide
Mindframe	Various guides for safe and accurate reporting and communication (primarily for media professionals and communicators).	Guidelines on reporting and portrayal of eating disorders Reporting suicide and mental ill-health Guidelines for image use Guidelines for language use

More information and online resources		
ReachOut	Provides a range of mental health and wellbeing information and resources for Australian young people. This includes a list of professionally reviewed mobile apps and tools available to support the mental health and wellbeing of young people.	Website Tools and apps
Kids Helpline	Provides a range of mental health and wellbeing resources and support for children and young people.	Website
headspace	Provides a range of mental health and wellbeing resources and support for young people.	Website
Beyond Blue	Provides a range of mental health and wellbeing support services and resources.	Website
Butterfly Foundation	Provides support and resources for people experiencing eating disorders or body image issues, and those that care for them.	Website
National Eating Disorders Collaboration	Provides a range of resources and tools to help you find the right service for you or someone you know who is experiencing concerns related to body image, food or eating. This includes a service locator tool which provides information about eating disorder-specific clinical services.	Website
Eating Disorder Families Australia	Provides a range of support and resources for carers and families impacted by an eating disorder.	Website