



Online Hate
Prevention Institute

Online Hate Prevention Institute submission on Modernising and Harmonising Classification Standards



Submission to the Department of Infrastructure, Transport,
Regional Development, Communications, Sport and the Arts

May 2026

About this submission

This submission is from the Online Hate Prevention Institute, an Australian Harm Prevention Charity established in January 2012 with a focus on online hate and extremism. This submission is prepared by Dr Patrick Scolyer-Gray, Dr Andre Oboler, Rachel Kaye and Jasmine Beinart, with reflections on the consultation documents “Update to modernise and harmonise classification guidelines 2025”.

Key Reflections

We agree that Australia’s classification guidelines have not kept pace with modern media trends, particularly gaming, streaming, and interactive environments. Each of these present unique challenges.

We are concerned that the consultation paper ignores the most important classification in the scheme: the Refused Classification (RC) rating. The existence of RC, the fact content can be refused another rating, what such a rating means legally, and what it means for the scheme as a whole is critically important.

We agree that the scheme should be designed to “protect without overreach”, however, the protection must be real which means the scheme must be more than a mere labelling scheme. We also agree that it should “reflect Australia’s diversity”, but this also means respecting common minimal standards which may at times be more important than platforming certain segments of society, for example, there are clearly neo-Nazis in Australia but that should not mean antisemitic and pro-Nazi propaganda films like “Europa: The Last Battle” should be permitted in the name of diversity. The same argument applies to videos produced by listed terrorist groups, even if there are supporters of those groups in Australia.

We strongly disagree with the idea that the scheme should “inform without censorship”. That is only true to a point. Beyond that point the scheme can and should act as censorship. This includes age based restrictions such as 18+. Protection also requires disallowing certain content entirely, which is correctly described as censorship. Content that parliament has deemed unlawful in Australia is subject to censorship, via a “Refuse Classification” (RC). This is a critical part of the classification scheme. One of the aims of the classification scheme system is to identify such content so that the marketplace and industry knows the content is prohibited and can voluntarily take appropriate action. Another is so that enforcement agencies have a clear mandate to take action when voluntary action is not taken.

On Refused Classification as part of the Scheme

Under the current classification system, abhorrent violent content (e.g. videos of live streams of terrorist attacks) and terrorist manifestos are required to be given an RC rating. Being able to point to the listing describing an RC item is useful in our work, for example when we approach platforms notifying them of such content and asking them to remove it. Where the content is not listed, our work is more difficult and the content can spread further, increasing

the risk of radicalisation and future deadly harm. This is discussed in our report on the Yom Kippur terrorist attack in Halle, Germany.¹ Content that is listed should be described sufficiently well to be identified. Some of the existing RC labels that have come from eSafety and provide only reference numbers are not sufficient to fulfil this purpose.

We also believe the action by the Classification Board in response to concerns raised by a women's advocacy group, Collective Shout, on certain computer games being sold without a rating, and which would likely be given an RC rating if evaluated, was appropriate.² The problem with large volumes of games promoting child exploitation, violent rape, etc. avoiding classification and being sold into Australia is not a problem of the classification scheme, it is a problem of the platforms selling such games. In our view, it may be appropriate to classify these games as RC, which could help prevent their sale in Australia. The Australian and Canadian schemes are, in our view, better in this respect than what occurred in the UK, where the content was only removed after protracted civil society campaigning with the support of a relevant government minister. This situation around harmful games, and the attacks on Collective Shout and their staff for highlighting unlawful content, is discussed in a major report that we will be releasing on August 12th at Parliament House, Canberra. We would welcome the attendance of the Department and the Minister at that launch.

On Reform Priorities and Appropriate Adaptation

We support the “shared priorities for reform” as listed. We are concerned by the reflection from industry participants about “inconsistent processes that slow classification and cause confusion” and the reference to the *Online Safety Act 2021* as the problem. The Online Safety Act is the most recent and modern relevant legislation and also something that has been heavily and frequently reviewed. We believe it is well adapted to the modern environment, and while that might mean a greater burden on industry, it is a burden in keeping with the greater financial scale, volume of content, cheaper production costs, and higher profits in this domain.

While we agree the public facing rating should be simple and consistent across mediums, the considerations which go into providing a rating may not be the same. The principle of *appropriate adaptation* means that a change in medium, which results in a change in the risk of harm, may justify a change in the legal response,³ or in this case the rating. To give an example, some books are known to be key sources that spread hate against particular communities and have a historical connection to incitement to violence leading to massacres. Two examples are the *Protocols of the Elders of Zion* and *Mein Kampf*. They are available in Australia as printed books. Whether they should be made freely available as online downloads or audio books is another matter. A further consideration is whether social media content that is designed to help the content spread and go viral is appropriate. The

¹ Andre Oboler, William Allington and Patrick Scolyer-Gray, *Hate and violent extremism from an online sub-culture : the Yom Kippur terrorist attack in Halle, Germany* (Online Hate Prevention Institute, 2019) <https://nla.gov.au/nla.obj-2286730824/view>

² Andre Oboler, “Gamer Gate Returns: Collective Shout edition”, *Online Hate Prevention Institute*, 9 August 2025. <https://ohpi.org.au/gamer-gate-returns-collective-shout-edition/>

³ Andre Oboler, “Legal Doctrines Applied to Online Hate Speech” (2014) vol 4 *Computers & Law* 9–15. <https://www.austlii.edu.au/cgi-bin/viewdoc/au/journals/ANZCompuLawJI/2014/4.html>

mediums can be seen as moving down a continuum of increasing risk of incitement and harm, as well as potentially removing mitigating context.⁴

While we generally support “processes that are inclusive and appropriately responsive to community sensitivities – issues such as discrimination, cultural harm and mental health” we are concerned by evidence before the Royal Commission into Antisemitism and Social Cohesion about the large problems of antisemitism and exclusion of Jewish people in the arts, including authors. Such systems can be abused. We would seek reassurances that the process will not lead to differential treatment of content based on the ethnic or religious identity of the content producer. The classification system processes must have protections against this.

On Aligning the Classifications Guidelines with Harm Prevention and Reduction Objectives for Children

The Classifications Board and the classifications system serve multiple purposes; one of which is the protection of people under the age of 18 (i.e., ‘children’) from harms associated with the consumption of certain material. Reducing harm depends on curtailing its reproduction and circulation across and between generations. Therefore, the OHPI strongly recommends that updates to the classification system should prioritise preventative controls and assign the responsibility for ensuring the safety of children to parents, guardians, and the government, as an intraoperative whole.

1. The classifications system must be built on the assumption that it is being used by parents and guardians to inform their own consumption choices *as well as* those for whom they are responsible. Harm prevention and risk aversion are often not factored into the decision-making of children (i.e., persons under the age of 18), which means that the responsibility of minimising the harm involved with children’s consumption of media cannot be assigned to children.
 - a. Although it is positive that some underage consumers find classifications ‘helpful’, we must still assume that the decision-making of children in terms of self-selecting media to consume when unsupervised involves far greater risk than it does when children’s consumption of media is monitored and moderated by a parent or guardian, or controlled by an age verification barrier.
 - b. We cannot entrust children to self-select media in accordance with the classifications system except insofar as explicit restrictions barring access based on age-verification are applied.
 - c. Therefore, as far as children are concerned, the **major** harm-reduction/prevention function of classifications applied to media lies in their capacity to inform the decision-making of adults (in lieu of their access being barred based on age-verifications).

⁴ “Mein Kampf on Sale”, *Online Hate Prevention Institute*, 18 March 2020. <https://ohpi.org.au/mein-kampf-on-sale/>; “Mein Kampf Discussion”, *Online Hate Prevention Institute*, 26 March 2020. <https://ohpi.org.au/mein-kampf-discussion/>

2. The core test for determining the classification of an item that cannot be classified as suitable for children must be based on the risks of harm that will arise in the event that a child selects to consume that media unsupervised.
 - a. The primary responsibility for the risk management of media consumption by children is parents and guardians, but we also know that:
 - i. Some children will not necessarily receive care and attention from parents or guardians with respect to the control of media consumption.
 - ii. Parents and guardians will necessarily make different assessments of the media that they believe is suitable for children to consume, and this variance must be accounted for in the development of a risk management strategy.
 - iii. Some children will, with varying frequency and degrees, circumvent the risk management strategies for media consumption imposed on them and still self-select and consume media regardless of warning labels and boundaries set by parents and guardians.
 - b. Therefore, residual risk is inherent to all classifiable content to which children can still access independently of an adult gatekeeper.
 - c. Recognising the points outlined previously [see 2.a.i-iii], It must be assumed that children will, on occasion, choose to consume any content regardless of classification except and unless access is precluded by age-gating (i.e., R or X ratings).
 - d. Therefore, all content that is assigned *any* classification that is neither G nor R18+ (or, obviously, X or RC) *must* be of a nature that an unsupervised child *could* consume without attracting risks of harm of the kind or severity that should be controlled by the R18+ rating.
3. The 'PG', 'M' and 'MA' ratings, as they are currently structured, are not as effective at reducing risk for underage consumers or adults as they could or should be.
 - a. Intermediate classifications do not perform a harm prevention/reduction role for adult consumers because the age 'suggestions' are irrelevant to adults; intermediate classifications warn about material that could be harmful to children, and therefore do not imply any such risk for adults. Adults should therefore not be exposed to any risk of harm when consuming any content that has been assigned any intermediate classification (i.e., if not rated R or X, media should never contain restricted or controlled content).
 - b. The age suggestions for intermediate classifications are not currently enforced commercially or domestically, which means that they do not prevent children from self-selecting material with those classifications [see C05].
 - c. In effect, the only remaining purpose of intermediate classifications is to assist adult consumers in their decision-making on behalf of underage consumers.
 - d. In this respect, intermediate classifications do not presently perform their function with consistency or accuracy, and must be updated to deliver what is, essentially, a summative representation of all the risks identified in a classified item.
4. To meet the harm prevention and reduction objectives of the classification system, the OHPI recommends that all classifiable content should be assigned a classification based exclusively on the outcome of a risk assessment that is both comprehensive in terms of the classifiable content considered, and inclusive of consideration of the severity rating of each classifiable sub-element.

- a. This means that the assignment of a classification rating to an item must be determined solely by the level of risk associated with the consumption of a classifiable item by an unsupervised child the same as outlined previously [see 2.d].
5. The function of the classification labels that are familiar to the Australian public, (specifically, PG, M and MA) can be retained, but their meaning and function needs to be clearly defined, and consumers must be able to understand that the pre-existing classifications system does not stand for or represent anything other than a corresponding quantity of risk that has been calculated based on the quantum of sub elements contained within that classified item.
6. To achieve these desired outcomes, the OHPI strongly recommends that the Classifications Board concentrate primarily on implementing Recommendations C02 and C03 with a general focus on enhancing the accuracy and completeness of sub-element categories as well as the implementation of consistent mechanisms for evaluating the severity of sub-elements.
7. By clearly defining classifiable sub-elements and evaluative schema for severity, the presence of risks to consumers should ultimately facilitate 'objective' and (crucially) *replicable* classification outcomes.
 - a. This will relieve the anxieties reported by content producers and industry about the threat of biases on the part of the Classification Board [see C12].
 - b. A uniform and replicable risk assessment system should also permit self-regulation and enhance the transparency of classification decisions.
8. The outcome of risk assessments of classifiable media, which must be based on the sum total of the risks arising from identified sub themes and their respective severity, can then be mapped to the existing classification system.
 - a. Mapping risk assessment outcomes to the PG, M and MA classifications is to be done to preserve the limited insight that they currently afford, but will now correspond with precision to a uniform continuum of harm and risk.
 - b. To preserve the pre-existing classification system, it will be necessary to determine three risk thresholds that, if reached, will then correspond to increasingly restrictive ratings.
 - c. By using the results of risk assessments and uniform risk thresholds, the heuristic afforded by the classification system remains accessible to parents and care givers that do not engage with the classification sub elements but still require a general point of reference in moderating the consumption of media by children in their care.
9. The introduction of the proposed risk assessment process, while it can still be mapped to existing classifications, will necessitate some clarification of what those classifications now mean.
 - a. 'PG' stands for 'Parental Guidance', which aligns with the purpose of the intermediary classification. Namely, risk assessments that produce scores that match the PG threshold should always be assigned to content that can be consumed by children, but that carries a degree of risk that also justifies risk assessments on a case by case basis by parents and guardians.
 - i. Following this system, a glance at a classification rating of PG will inform all consumers that the content:

1. Carries a risk of harm to some degree if consumed, but that whatever that extent might be, it is not so severe as to necessitate a classification of 'M'.
 2. Is more likely than not to be suitable for a person aged 15 or older to safely consume without excessive risk of harm.
 3. Should be evaluated by a parent or guardian before being consumed by a person under the age of 18.
 - b. 'M' stands for 'Mature', and will correspond to items that achieve a score that meets the next risk threshold above that of PG.
 - i. Following this system, a glance at a classification rating of M will inform all consumers that the content:
 1. Carries a risk of harm to some degree if consumed, but that whatever that extent might be, it is not so severe as to necessitate a classification of 'MA'.
 2. Is of a nature that a person under the age of 15/16 has significant risk of being harmed as a result of consuming the content.
 3. Should be evaluated by a parent or guardian before being consumed by a person under the age of 18.
 - c. 'MA' currently stands for 'Mature Accompanied', but in light of the practical problems and inconsistencies in the enforcement of the current definition [see C05], this will need to change to the equivalent of 'restricted to mature audiences of persons 15 (or 16) and older'.
 - i. Following this system, a glance at a classification rating of MA will inform all consumers that the content:
 1. Carries a risk of harm to some degree if consumed, but that whatever that extent might be, it is not so severe as to necessitate a classification of 'R'.
 2. Is of a nature that, if consumed by a person younger than 15/16, is significant risk of causing harm.
 3. Is restricted to persons over the age of 15/16.
 - d. As previously discussed, *any* content that presents a risk of harm that cannot be tolerated as acceptable for a person under the age of 18 must always be classified as R. However, this classification must still only ever be assigned to items that meet the risk threshold following the same standardised risk assessment.
10. The OHPI also notes that this proposed adjustment will necessitate re-classification of existing items currently classified as PG, M or MA via the new system based on scores of risk derived from the presence of sub elements and their respective severity.
- a. Such an undertaking aligns with the reported interest in overhauling the classification sub-elements [C02] and their respective severity ratings [C03] to achieve a greater level of consistency across all classified material [C07].
 - b. Updates to classifications would not need to be immediate, but could instead be rolled out gradually.
 - c. Alternatively, the responsibility for reclassifying content could be transferred to content providers, who would then face a financial incentive to self-regulate

- and classify content in accordance with the Classifications Board system to maintain access to Australian markets.
- d. Items classified as R under the previous standard could also apply for evaluation under the new system if it is thought that doing so might result in a different rating.
 - e. It is expected that classifications for a significant proportion of material currently classified as M or MA will, under the proposed risk assessment strategy, become re-classified as MA or R. This is because the revised baseline risk appetite outlined within these recommendations are significantly reduced [see 2.d].
11. In accordance with the stated interest in incentivising content producers and industry to meet Australian classification standards autonomously and without the need for direct oversight by the Classification Board, the proposed change to the classification system would significantly alter (and in many ways reduce) the complexity and burden of compliance.
- a. Content producers seeking to distribute media to Australian consumers could rely on a comparatively simple risk assessment matrix to self-regulate and to assess their compliance.
 - b. It is assumed that the sales of content previously classified as M or MA are unlikely to be harmed by the proposed change, and neither too should sales of content rated R (prior or following the recommended change's implementation). This is because the existing market share for material classified as R should be completely unaffected (given that consumers of such material should be over 18 anyway), and re-classifications from M and/or MA would, in theory, serve only to *increase* the potential market share.
 - c. The burden for industry in terms of meeting Australian classification standards would become far more predictable because of the central and deciding role performed by sub elements.

Proposed Reforms and Next Steps

The "Classification Advisory Council" to provide ongoing community input is useful, but in light of the increased focus on discrimination, cultural harm and mental health, relevant organisations need to be given some form of standing to bring concerns to this body or to be consulted when issues relevant to them come up for discussion.

Regarding "stronger regulatory oversight of growing industry self-classification and automation tools to improve efficiency while maintaining accountability," it is our view that industry, particularly the social media industry, has in the last 18 months taken a huge step backwards in self-regulation. We believe self-regulation is failing. We also fundamentally believe that automation is problematic in this space. The automation can only spot what it is trained to spot, which means it can say when content may be problematic in a way it recognises; it cannot identify problematic material that it has not been trained to spot. It therefore fundamentally cannot say that something is free of concerns - that requires human judgement. Too much reliance on automation results in a 'publish first, then address the damage if the fall out is bad enough' mentality. That is the model social media has used, and which has led to an increase in harms in society. It was somewhat mitigated by automated responses that removed a significant amount of harmful content before anyone saw it, but

those systems have, on Facebook and Instagram for example, been turned off, leading to a significant reduction in the amount of harmful material that is being removed and a corresponding increase in the level of public harm.⁵ We should not spread this problem through the whole system.

⁵ Exhibit 1-116 – STAT.8999.0057.0022_R – Dr Andre Oboler, *Royal Commission into Antisemitism and Social Cohesion*, paragraph 98
<https://asc.royalcommission.gov.au/publications/exhibit-1-116-stat899900570022-r-dr-andre-oboler>

Responses to the Recommendations

#	Element	Description
C01	PG13 Rating	The guidelines could be updated with a new classification rating - PG13.
C01 - Response: If intermediary classifications must be retained, then they should not be updated or modified. Doing so will contradict the need for streamlining and compliance by self-regulating/classifying content producers as specified in Recommendation C12.		
C02	Overhaul of classifiable element 'Themes'	Themes could be re-worked to clearly classify, identify and communicate specific community needs so that the resultant classification truly reflects the content. Genre descriptors (e.g. science fiction, fantasy) could be removed from Themes. Themes would act as a collection of sub-elements for classification and reflect the most requested issues. These would be: Sexual violence, suicide, discrimination and scariness. When a classification matrix for a title is shown, the sub-matrix of Themes would also be visible.
C02 - Response: - It is crucial that the overhaul of sub elements includes the addition of severity ratings for each element [see C03]. - Additional sub elements that need to be afforded equal (and sometimes greater) weight than those mentioned already must also include racism, misogyny, antisemitism, xenophobia, islamaphobia, homophobia and any other prejudicial, discriminatory and hateful ideologies, beliefs and associated symbols, practices and any other related materials. 'Incitement', particularly to violence, should also be included as a prohibited sub-element, and media containing it should be refused classification. - The Classifications Board, when overhauling or updating classifiable elements, should seek direct consultation and support from organisations such as OHPI to establish effective working definitions of specialised subject matter (e.g., antisemitism, misogyny, xenophobia etc.). For example, OHPI has established a schema of anti-Asian hate which may be of assistance to the Classification Board.⁶ - Media that is identified as hate speech should be refused classification. - Content that prominently features Holocaust denial material should may be refused classification.⁷ - Content that <i>promotes</i> and/or is <i>instructive</i> in sexual exploitation and/or misogyny must be refused classification.		

⁶ Andre Oboler, *Anti-Asian Racism in Australian Social Media* (Online Hate Prevention Institute, 2022), <https://ohpi.org.au/anti-asian-racism-in-australian-social-media/>

⁷ Andre Oboler et al. *Online Antisemitism After 7 October 2023* (Online Hate Prevention Institute, 2024), <https://ohpi.org.au/afteroct7/>

		For example, some of the productions of so-called 'pickup gurus' should be refused classification.
C03	Improved reliance on evidence over morality	Guidelines would be updated to clearly define terms and thresholds to balance both community concern for certain content with any established evidence-based research of actual harms.
C03 - Response: - Eliminate 'community concern' - such 'concern' likely to be based on a subjective assumption/belief/predisposition, which contradicts the stated objective of Recommendation C03 to being evidence-based. - Definitions and thresholds, once defined and adopted by the board, should only be subject to change in the event of a successful formal application. - The revised thresholds and definitions for classifiable content must concentrate on clearly and consistently evaluating the severity of sub elements identified in classifiable material. - The OHPI has not defined the scales for rating the severity of sub-elements, but it recommends that they be based on the extent to which the sub-element presents a risk of harm to a person of the minimum age to which the 'PG' rating is applicable. - This is because the primary risk of harm from the consumption of media is to children, and to accommodate for variance and to minimise risk, a conservative approach is necessary. By selecting the youngest aged person who could have access to material classified PG, we can ensure that we use a universal benchmark for evaluation of severity that also caters to the most vulnerable consumers (and in so doing, also necessarily caters to older and ostensibly less vulnerable consumers). - Scales of severity should be based on multiple point scales of low complexity that assign the lowest value to mean 'harm that is so low in severity it cannot be classified any higher', and the highest value should correspond to mean 'the potential harm presented by this sub-element is as high as it can be without crossing the boundary into necessitating an R classification'. - For example, the severity system could be as follows: 1 - The sub-element presents the least risk of harm, but still presents some risk. 2 - The sub-element is of a severity that some children are at risk of being harmed by exposure. 3 - The sub-element is of a severity that most children are at risk of being harmed by exposure. 4 - The sub-element is of a severity that most children are at high risk of being harmed by exposure. 5 - The sub-element is of a severity that is as high as it can be without crossing the boundary into necessitating an R classification. - Additional steps to calibrate classifications could include simple measures such as considering if a classification for a sub-element matches the logical placement of it on a scale, e.g., is this more severe than one classification lower than it is, and is it also less severe than one classification higher than it presently is?		
C04	Interactivity classifications focus on replication	Currently, classifications of elements in video games have a stricter benchmark when interactivity is involved. If a video game includes violence or nudity at the same level as a film, the game will be considered more extreme. The classification guidelines could be updated to remove the harsher judgement of video games' interactivity, except when it comes to

		real-world replication. E.g. violence in a video game is not the same as real world violence, but gambling in a video game for a reward or incentive is, in effect, the same as gambling in the real-world.[PS2]
C04 - Response: • Research clearly indicates that interactivity is just as if not more influential on behavior than the consumption of films. Therefore, the existing benchmark, which is more strict for videogames, should be preserved. • It is a mistake to presuppose that violence in video games requires less regulation than gambling in video games - they have different risks, are categorically different, and should not be compared. • Given that the research conducted for these proposed changes found that “only one-third of those aged 13-17 said that seeing a gambling warning would make them less likely to play these games”, it is clear that the function of classifications is to assist parents and guardians, not children. As discussed previously in this submission, it is reiterated here that if interactivity sub elements are to be included in a classification of risk, then they should be considered and published in a manner primarily designed to facilitate informed consumption choices by adults on behalf of children. If the interactivity is of such a concern and presents sufficient risk of harm to make the prospect of a young child consuming such media unacceptable, then it should be classified as R. • The OHPI recommends that interactivity in video games that presents foreseeable risk of self-radicalisation should be prioritised as prohibited or restricted content, with careful consideration given to the possible need for such content to be refused classification. • A major priority for the classification of video games in Australia is to ensure that they actually receive a classification. • Online game marketplaces are used to self-publish games, likely in numbers exceeding the capacity of such marketplaces to self-regulate (Steam and itch.io are two prime examples of international marketplaces). ○ Large and highly dynamic market places might be beyond the regulatory scope of the Classifications Board, and it might be unavoidable that self-regulation become the default requirement in such spaces. ○ However, if self-regulation is necessitated, then the reporting of content and referral for formal classification of content found to be in breach of Australian classification standards must be accessible to the general public. ■ The resources required for a manual review of content can be recovered from content producers, platforms and/or service providers by way of fines/fees.		
C05	MA15→ MA16+	The MA15+ classification could be updated to MA16+, reflecting the social benchmark of age sixteen, and other legislation such as the Online Safety Act that prohibits certain online environments for those under sixteen.

C05 - Response:

- The MA classification is an intermediary classification that adds complexity to the classification process, and in so doing, contradicts the stated desire to rely on industry self-regulation (C12).
- If the MA16+ classification is to be implemented, it will only serve its intended purpose if consistently and universally enforced via age verification systems.⁸
- If retained, the MA16+ classification's implementation will need to accord with the overarching logic of the classification system. Specifically, the content classified as MA16+ must consistently be of a nature that is more harmful than content assigned a rating above G, but less harmful than content assigned a rating of R.
 - Notably, this is essentially identical to the definition of the proposed 'PG' classification, and therefore introduces ambiguity and risks confusion. This contradicts C07.
- The MA rating is premised on a parent or guardian being in attendance for the duration of consumption of all media assigned this classification. This requirement might be practical for films or TV, but it is an immensely high resource cost to expect guardians to attend for the duration of gaming sessions or incidental consumption of material that is available via mobile devices. To maintain consistency between mediums, the classification needs to abandon the 'must be accompanied' component before it is practically possible to enforce.
- If a classification cannot be enforced reliably, then it will become used as a means of avoiding higher or lower classifications by content producers, and that will undermine the integrity of the entire classifications system.

C06

Publications to adopt the rest of the classification system

Publications currently have a unique classification system unlike other media - Unrestricted, Unrestricted Mature, Category 1 Restricted and Category 2 Restricted.

To simplify, publications could adopt the same classification system, but only using the relevant levels:

Unrestricted - General

Unrestricted Mature - MA16+

Category 1 Restricted - R18+

Category 2 Restricted - X18+

C06 - Response:

- This recommendation appears to be acceptable, but additional consideration of the unique power dynamics/authority afforded non-fiction is needed.
 - Similar to the 'replicability' concern of C04, some non-fiction content promotes or instructs the consumer in dangerous or harmful behaviours.
- Additionally, and as demonstrated in examples of material such as *Mein Kampf*, serious consideration of a review of existing classifications for extremist materials should be given, with a focus on the assignment of R, X and RC.
- It is essential that terrorist materials, like manifestos and videos, also receive an "RC" rating to ensure they are not spread in Australia through other offline formats.⁹

⁸ Andre Oboler, *Aboriginal Memes & Online Hate* (Online Hate Prevention Institute, 2012), <https://ohpi.org.au/aboriginal-memes-and-online-hate/>; Andre Oboler, *Submission to the Joint Select Committee on Social Media and Australian Society* (Online Hate Prevention Institute, 2012), <https://ohpi.org.au/ohpi-submission-on-online-safety/>

⁹ Andre Oboler, David Wishart and Mark Civitella *Submission to the Consultation on a Bill for a New Online Safety Act: Department of Infrastructure, Transport, Regional Development And Communications* (Online Hate Prevention Institute, 2020)

C07	Clarify definitions	The current language in the guidelines is ambiguous, which can lead to broad interpretation and potential bias in classification decisions, depending on how instructions and definitions are understood. To address this, clear and well-informed definitions and instructions should be developed alongside updated evidence, ensuring the guidelines provide objective and understandable direction for classifying content.
-----	---------------------	---

C07 - Response:

- The OHPI concurs with and endorses this recommendation, and emphasises that the criteria and definitions that guide classification must be based on harm-prevention/reduction objectives and criteria.
- The range of sub-elements and severity scales must be comprehensive and clearly defined as per Recommendations C02 and C03.

C08	Reconsideration of sexual content	Currently, sexual scenes involving participants that are non-heteronormative (a man and a woman) are often automatically considered higher impact, as are depictions that include fetish or kink-related acts typically judged through a moral lens rather than an evidence-based, harm-informed framework. Except in cases involving depictions of sexual violence or non-consensual acts, sexual content should be rated consistently, regardless of the artistic style or identities of those involved.
-----	-----------------------------------	--

C08 - Response:

- The OHPI concurs with this recommendation insofar as equal treatment of consensual, non-violent sexual content regardless of the subjects' gender identity, sexuality and/or preconceived notions of deviance and taboos etc.
- However, the OHPI will also emphasise that current research clearly shows that the exposure of sexual content of *any* kind to children incurs a risk of harm. The more severe the sexual content, the greater the harm.
- In light of Australia's ongoing problems with structurally entrenched social issues such as family violence, sexual violence and misogyny, the OHPI recommends that media containing sexual content should, wherever possible, be classified with special consideration, and with a general objective of restricting consumption by children.
 - However, the OHPI will also reiterate the central criticality of preserving individual liberties and freedom of expression. The classification of sexual sub elements *must* be based on a risk assessment of harm, with due consideration afforded to the corresponding need to avoid replicating the puritanical censorship regimes past and present.

C09	Classification Advisory Council	A Classification Advisory Council, composed of community voices, not professional industry stakeholders, could be created to not only aid in this process, but also provide a productive method by which industry and government can gain feedback on decisions and ideas.
-----	---------------------------------	--

C09 - Response:

- The creation of an advisory council may add administrative overhead and complexity to the classification process, which contradicts the objectives of C12.
- There is a significant risk of introducing lobbying and the individual biases identified as concerns in C12 into the classification process by opening participation in the classification process indiscriminately. This risk also contradicts recommendation C03.

- However, as per previous OHPI recommendations, outsourcing of some evaluative and adjudication functions to non-government organisations could assist and streamline the classification process.
- Therefore, the OHPI recommends the creation of an advisory council to support and facilitate the Classifications Board so long as it is composed **exclusively** of accredited or otherwise suitably experienced professionals.
- Those undertaking the review of items by and/or for the Classifications Board should receive specialised training in the identification of hateful content, coded language and other information relevant to effective classification.

C10	Single touch Classification	Titles could be classified once, only to be updated again if the content itself changes. This would mean if a film is classified for theatres, it automatically carries the same rating for streaming, etc. This would reduce the burden for industry and improve the experience for consumers.
-----	-----------------------------	---

C10 - Response:

- The subject matter of material is what needs to be classified. As such, there is no reason for content that is transferred from one medium to another to require additional classification. The exception to this would be instances where the content's capacity to inflict harm is altered as a result of the change in medium (e.g., from a book to a film, a film into a video game etc.).
- Equally, the OHPI concurs that arbitrary repetitions of the classification process triggered by inflexible and illogical administrative systems should be avoided.

C11	Permit updates to classifications	Currently, any change to a piece of media requires the entire content to be reclassified. Instead, the database could allow for classifications to be updated. A record of the previous classification would be retained on the listing - if the classification changes.
-----	-----------------------------------	---

C11 - Response:

- The capacity to update classifications must be retained, and given the centrality of harm-prevention/reduction, a conservative approach is warranted. Specifically, updates to classifications should remain at the most restricted level of classification assigned to that item unless the Classification Board can be satisfied that all previous versions of the piece of media have been adapted to meet the requirements of a new classification.
- Notification of changes to classifications should be automatically distributed to content producers (i.e., publishers, authors, providers etc.), along with information about the content producer's corresponding regulatory obligations. This should help facilitate swift and efficacious compliance, such as through content updates, retractions, suspensions etc.
- Updates to classifications may prove particularly important in cases where the emergence of new information necessitates updates to media that contains materials which, as a result of emerging information, might be found to contain mis/disinformation.
- A log of changes to the classification of every item should be maintained as a standard requirement of information management and as a prerequisite to the integrity of the Classification Board as well as its accountability.

C12	Self-classification as default	A significant proportion of content is still classified by the Classification Board—a process that is both time-consuming and costly. There is also concern from industry - whether factual or not - that outcomes can be influenced by individual biases, depending on the composition of the Board. Rather than relying on Board classification, self-classification could become the industry standard, supported by evidence-based tools - with Board classification only in rare circumstances.
-----	--------------------------------	--

C12 - Response:

- Defaulting to self-classification by industry is an attractive idea, but the OHPI have also observed that such initiatives are unreliable, and non-compliance becomes increasingly difficult to remedy the more that responsibility for compliance is privatised.
- If a default of self-classification is pursued, then content providers will need support to assist them in meeting their compliance obligations with Australian classification standards.
 - Given the expertise required to identify and appropriately manage problematic and classifiable sub elements such as antisemitism and misogyny, the implementation of regulatory compliance may create a market for consulting and advisory services.
- Any attempt to default to self-classification will need to be supported by a robust and comprehensive system for enforcement of Australian classification standards. These will likely need to be a combination of market restrictions and fines, as well as remedies for more severe transgressions such as takedown powers.
- In general, the OHPI recommends against relying on industry to self-regulate, but there are some exceptions to this rule.
 - Specifically, dynamic content, such as participatory environments that host UGC, live streams, and other sources such as online media marketplaces, produce significantly more content than can reasonably be regulated manually by the classification board.
 - Therefore, the OHPI recommends establishing a system to support compliance by content producers in tandem with community reporting [see C04].

C13	Modernise the classification database	Updating the database so titles are nested together, potentially utilising AI/natural language to aid in search functionality to remove potential error from different spelling/grammar, and the reduction of required elements for an entry could aid in making the user experience of the database more manageable.
-----	---------------------------------------	---

C13 - Response:

- OHPI has no comment on this recommendation.

C14	Expand use of automated classification tools	Expanding the use of automation and self-classification tools could support industry to meet classification obligations more efficiently. This may involve reviewing existing tools, introducing new ones and investing in improved technologies aligned with updated definitions. These tools would not only streamline the classification process but also enable industry to more effectively reach their intended audiences.
-----	--	--

C14 - Response:

- OHPI's observations and research evidences considerable risk in reliance on automation and AI tooling for the classification of content in accordance with harm-reduction/harm-prevention objectives.
- Defined decision-trees for the assessment of sub-elements might be automated if they can interface with machine readable renderings of the data (e.g., word and phrase searches could be performed automatically if the script of a film's dialogue is provided).
 - However, even such basic approaches are still vulnerable to errors, such as the use of coded language, metaphors and idioms, as well as the context in which language is embedded (e.g., a script may never use the word 'Nazi', and thus avoid classification for such content despite characters depicted in Nazi uniforms). There is significant risk that automated tools will miss significant amounts of potentially harmful content.
- Automated assessments of reportable content in environments like social media provide a compelling point of comparison to other sources of classifiable media, such as UGC and live streaming.¹⁰ OHPI's research has shown an exceptionally high failure rate by automated tools in their ability to reliably recognise anti-social content reported on social media platforms.¹¹ It is unlikely that content producers will consistently and accurately classify their own content if relying primarily on AI/automation.
- The purpose of the classification is to identify and control the risks to which consumers are exposed, not to reduce the risks of content, which would necessitate altering the subjects of classification. If the proposal is for content producers to invest in tooling that modifies their content to accord with Australian classification standards, then this should be a matter left to the exclusive purview of content producers. The Classifications Board must *never* attempt to alter or modify content, and it must not impose barriers to entry through preferential treatment to content producers that are able to invest in tooling that is not universally accessible.

C15	Classification website - uniform matrix	The website could be updated and refreshed to always show (if the data is available in the current guideline era) the classification matrix, ensuring consistency and usability for consumers wanting to know more about certain content.
-----	---	---

C15 - Response:

- This step is essential for the sake of fulfilling automation and AI objectives as specified in C14, and should be prerequisites to the transparency recommendation of C18.
- The OHPI concurs with this recommendation, and emphasises how the detail and accuracy of the matrix should accord with the updated classifiable sub elements of Recommendation C02.

C16	Classification website - personalised preferences*	The classification website could allow users to create accounts and set personalised preferences for themselves or those they care for. When logged in and viewing a title, a consumer could then see how this title aligns with their chosen settings. It could also direct the consumer to
-----	--	--

¹⁰ Andre Oboler, 'Poway Synagogue Attack', *Online Hate Prevention Institute*, 28 April, 2019. <https://ohpi.org.au/poway-synagogue-attack/>; Andre Oboler, William Allington and Patrick Scolyer-Gray, *Hate and violent extremism from an online sub-culture: the Yom Kippur terrorist attack in Halle, Germany* (Online Hate Prevention Institute, 2019) <https://nla.gov.au/nla.obj-2286730824/view>

¹¹ Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Analysis of the Online Abuse Experienced by Collective Shout 2025-2026* (Online Gate Prevention Institute, [Forthcoming]).

		important resources and information relevant to the content - e.g. how to introduce teenagers to higher rated content safely.
C16 - Response:		
The OHPI has no comments on this recommendation.		
C17	Classification website - unify listings	When viewing a listing on the classification website, it currently provides a list of all possible iterations or versions of one single title. The website could be updated to provide one listing for a title, for ease of search, and then include the ratings under this for different versions or formats. This might greatly improve the usability of the website and reduce confusion for consumers.
C17 - Response:		
The OHPI has no comment on this recommendation apart from emphasising the criticality of balancing ease of access to information with the need for accuracy and transparency. The convenience of the consumer is of less importance than the elimination of ambiguity and the supply of detailed information pre-requisite to efficacious decision-making.		
C18	Classification website - transparent decisions	The website could show a log of the questions, concerns and decisions raised/made by the Classification Board or Review Board for listings, especially where classifications were contentious. This communicates the consideration and effort undertaken by the classification process, so the public further understands the nuance that goes into classifications.
C18 - Response:		
- OHPI endorses the proposal for increased transparency of Classification Board decisions. - Recommendation C18 accords with the following OHPI recommendations: - The Classification Board should restore the previous tool that allowed more detailed interrogation of Classification Board decisions, specifically, it should allow all decisions in a given period for a particular classification, to be listed.¹² - The Classification Board should ensure either the title or a useful description is provided for material which is given an RC classification. This is necessary as the public cannot comply with a ban if the banned content cannot be identified.¹³ - A primary objective of the transparency described here should be to secure the trust of consumers and preserve the authority of the Classifications Board.		
C19	Classification website - addition of extra community information	The public actively seeks input and perspectives on content and classifications from other organisations and community groups. Further information and nuance are often sought to understand content appropriateness. Example resources/groups include Children & Media Australia, Common Sense Media etc. The website could display a summarised dashboard of community input from these groups on content listings and provide additional links to their resources as well as those of other organisations, such as eSafety.

¹² Andre Oboler, William Allington and Patrick Scolyer-Gray, *Hate and violent extremism from an online sub-culture : the Yom Kippur terrorist attack in Halle, Germany* (Online Hate Prevention Institute, 2019) <https://nla.gov.au/nla.obj-2286730824/view>

¹³ Ibid.

C19 - Response:

- Crowdsourced intelligence systems are frequently abused as a means of sanctioning, censoring and/or punishing targeted people and social groups.
- Any efforts to integrate processes of automatic consolidation of information from unregulated sources carries significant risk of harms that greatly outweigh the potential benefits afforded by convenience.
- It is recommended that classification markings include clear directions to consumers to online spaces in which they can find further information about the title should the consumer choose to do so.

C20	Element icons	Icons could be implemented to highlight the most important and highest rated classifiable elements (e.g. Nudity, Violence, Sex) in a given media. Specialised icons could be used to highlight areas of significant community concern when they are otherwise hidden under 'Themes' (e.g. sexual violence, suicide).
-----	---------------	--

C20 - Response:

- The classifiable elements should be listed, consistently, in descending order of the categories of harm that are featured with the most severity, rather than in accordance with a hierarchy of types of harm, which rank importance independently of severity of that type.
- Simplicity encourages accessibility, and as such, it is recommended that the resources required to design and implement a new system of symbols (as well as the associated risk of confusion stemming from variance in interpretive tendencies) be instead concentrated on maintaining easily understandable, accurate, comprehensive and consistent lists of classifiable elements, along with their respective severity/intensity/frequency ratings.
- Again, the highest priority in terms of the presentation of classification information is to ensure that parents and guardians are equipped to make appropriate choices when selecting media for children, and similarly, for children who are self-selecting media for consumption. Further investigation of the efficacy of iconography as a means of deterring children from self-selecting and consuming media without parental oversight may be worthwhile, but whatever is done, the central concern lies with facilitating informed consumption of media by children and parents/guardians.

C21	Quick links to classification listings	Classification titles could come with an option to quickly access the title listing on the classification website, via QR Codes (or similar technology) for easy scanning. This would mean that some information could be displayed firsthand, while deeper detail is readily available via a smartphone scan.
-----	--	---

C21 - Response:

- The OHPI concurs with the recommendation to make more detailed information about the classification of content available to consumers.
- Improving accessibility to relevant information is encouraged, but such efforts must not be pursued at the expense of safety or as increased costs to content producers. QR codes create a cybersecurity risk for consumers, and the maintenance of overly detailed databases of information about classifications may become prohibitive to the Classification Board and/or content producers.

C22	Generational education campaign	Misunderstandings about classifications has led to confusion about its function and place in society and the modern media landscape. A public campaign to remind Australians of the purpose and functionality of the classification system could be implemented. It could provide clear education and communication around its role in advice and restriction of certain content, that it is a point of reference for information so that individuals can make informed decisions for themselves and their children, and clearer understanding on what the symbols and levels mean.
-----	---------------------------------	---

C22 - Response:

- OHPI concurs with this recommendation.
- The fact that there has been any confusion in the first place lends support to the practical utility of reducing the complexity of the classification system by removing intermediary classifications.
- The ability of members of the general public to refer unclassified media to the Classifications Board for review should be promoted to the general public, and the process streamlined.¹⁴
 - This is particularly important for the purposes of maintaining consistent and up-to-date regulation of online content purveyors, such as online game marketplaces.

Additional Recommendations

1. Content that runs a significant risk of promoting or in any way facilitating self-radicalisation should be strongly considered to be refused classification, and in cases where this is not appropriate, it should likely be assigned an R18+ classification.
2. The classifications board's powers, especially as they interact with the eSafety commissioner's office, as a mechanism for assigning RC classifications to harmful online content, should be reinforced and highlighted as part of any revisions to the classifications system.¹⁵
3. To prevent attempts to circumvent restrictions on the possession and circulation of online content via transference to a physical medium, the Online Safety Act should be updated so that it can proscribe content that is stored on a physical medium. Determinations of the Commissioner should count as a classification, pending a classification from the Classification Board, or determinations should trigger a later classification by the Classification Board.¹⁶
4. Non-compliance with Classification Board determinations (e.g., refusing or failing to implement age verification requirements, update warnings and labels etc), where

¹⁴ Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Analysis of the Online Abuse Experienced by Collective Shout 2025-2026* (Online Gate Prevention Institute, 2026) [Forthcoming].

¹⁵ Andre Oboler, *Aboriginal Memes & Online Hate* (Online Hate Prevention Institute, 2012), <https://ohpi.org.au/aboriginal-memes-and-online-hate/>

¹⁶ Andre Oboler, 'Submission to the Inquiry into the Online Safety Bill 2021: Senate Standing Committee on Environment and Communications, Australian Parliament', *Online Hate Prevention Institute*, Jul 2024.

identified, must be prosecuted through the coordination of the Classification Board, eSafety, as well as relevant law enforcement agencies.¹⁷

5. Submissions to the Classification Board that, upon review, feature content that presents an adequately significant threat, should be refused classification, and the Classification Board should make referrals for takedown requests and removal orders to relevant authorities (e.g., eSafety, Australian Human Rights Commission etc.).¹⁸
6. Enforceable expectations should be set by government and exist outside of the terms of service of any one company, and should be consistent for different providers with respect to equivalent services.¹⁹
7. A description of proscribed abhorrent violent content should be maintained either with the Classifications Board or by eSafety, or both, in order to facilitate voluntary removal (potential on advice from online safety focused civil society organisations) by companies who find they are hosting such content.²⁰
8. Consideration of the possibility of expanding the remit of the Classifications Board to include review of certain broadcast media, including journalistic programming, should be undertaken in consultation with relevant experts.²¹
 - a. TV programming and news media can become a vehicle for hate speech as well as incitement. When such sub elements are identified in media, the content producers should be warned, and if they fail to self-regulate, their capacity to broadcast on Australian stations should be curtailed by the corresponding controls associated with an R, X or RC rating.
9. Chatbot/AI/LLM systems/services themselves should be assigned classifications. This aligns with the principles outlined in Recommendation C04, which highlight the risk of harms associated with interactive elements of certain media.
 - a. This serves to reduce the risks of the harms to which children using such tools are exposed.
 - b. Classifications and corresponding restrictions on market reach/share may also incentivise refinements of the tooling on the part of service providers.

Other comments

The Classifications Board must accord with a self-imposed mandate to preserve and protect freedoms of expression and speech as one of the highest (but certainly not sole) priority. In addition to the general social benefit of promoting and preserving expression, it is essential that this objective be made visible so that public support of the Classifications Board and its

¹⁷ Andre Oboler, 'Address to the United Nations Forum on Minority Issue', *Online Hate Prevention Institute*, 21 Nov 2021, <https://ohpi.org.au/address-to-the-united-nations/>

¹⁸ Andre Oboler, David Wishart and Mark Civitella *Submission to the Consultation on a Bill for a New Online Safety Act: Department of Infrastructure, Transport, Regional Development And Communications* (Online Hate Prevention Institute, 2020)

¹⁹ Andre Oboler, 'Online Hate Prevention Institute submission to the statutory review into the operation of the Online Safety Act 2021, Department of Infrastructure, Transport, Regional Development, Communications and the Arts, Australian Government', *Online Hate Prevention Institute*, Jul 2024.

²⁰ Andre Oboler, 'Online Hate Prevention Institute submission to the statutory review into the operation of the Online Safety Act 2021, Department of Infrastructure, Transport, Regional Development, Communications and the Arts, Australian Government', *Online Hate Prevention Institute*, Jul 2024.

²¹ Andre Oboler, *Measuring the Hate: The State of Antisemitism in Social Media* (Online Hate prevention Institute, 2016), p. 2, <https://ohpi.org.au/reports/Measuring%20the%20Hate.pdf>

authority are also protected. It is important that the Classifications Board does not become, and does not perceive to have become, an administrative appendage of unjust or arbitrary censorship. However, this directive must necessarily be balanced against the highest mandate, which is to reduce harm.