THE ONLINE HATE PREVENTION INSTITUTE

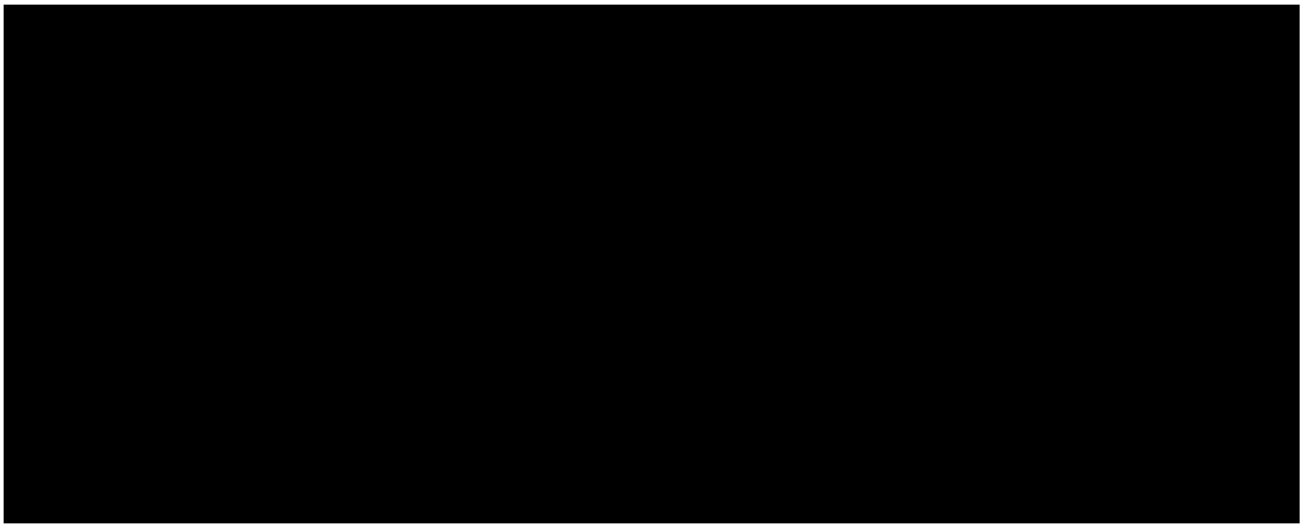
Empowering communities, organisations and agencies in the fight against hate.

ONLINE HATE PREVENTION INSTITUTE SUBMISSION

CONSULTATION ON A BILL FOR A NEW ONLINE SAFETY ACT

SUBMISSION TO THE DEPARTMENT OF INFRASTRUCTURE, TRANSPORT, REGIONAL DEVELOPMENT AND COMMUNICATIONS

14 FEBRUARY 2021



ABOUT THE ONLINE HATE PREVENTION INSTITUTE

The Online Hate Prevention Institute (OHPI) is Australia's only harm prevention charity dedicated to tackling online hate and extremism. It has been doing so since January 2012. We thank the Government for the opportunity to provide this submission.

Since our establishment in 2012, we have worked with various parts of Government to enhance online safety. We have worked with, *inter alia*, the Australian Federal Police ("AFP") and through them the Australian Intelligence Community, the Attorney General's Department ("AGD"), the Department of Foreign Affairs and Trade ("DFAT"), the Department of Veterans Affairs, the Australian Human Rights Commission ("AHRC"), and various state government departments and police forces.

OHPI has been involved in the consultations which led to the current online safety system, including the consultation on the Coalition's 2012 *Enhancing Online Safety for Children Discussion Paper* and the Government's 2014 *Enhancing Online Safety for Children* consultation, and the 2020 consultation on a New Online Safety Act.

We have engaged directly with eSafety, and regularly find ourselves at the same meetings and conferences, both within Australia and internationally, as eSafety staff, representatives from AFP, DFAT, and AGD. We are often the only Australian civil society group at these meetings, usually invited by overseas organisers who recognise us as a global leader in this space. It is a matter of regret that local organisers seldom engage with civil society. We acknowledge eSafety does engage to a limited extent with civil society stakeholders, albeit in very narrow areas. The importance of civil society to online safety is dealt with below.

Our focus on online hate and extremism covers hate against individuals (e.g. cyberbullying, serious trolling, RIP trolling etc), hate against specific groups within society (e.g. not only antisemitism, Islamophobia, homophobia, misogyny, racism, transphobia but also serious attacks on other groups such as ANZAC veterans and politicians, etc.), and hate targeting our society as a whole; for example, white supremacy and other form of violent extremism. In recent years, our focus on combating violent

extremism has increased as online radicalization has grown. We have been actively involved in removing terrorist manifestos and abhorrent violent content videos, and monitoring social media for threats.

ABOUT THE AUTHORS

DR ANDRE OBOLER

Dr Andre Oboler is the CEO & Managing Director of the Online Hate Prevention Institute. He is an Honorary Associate at La Trobe Law School, a global Vice President of the IEEE Computer Society, a member of the Global Public Policy Committee of the IEEE and an expert member of the Australian Government's Delegation to the International Holocaust Remembrance Alliance.

Andre was formerly a Senior Lecturer in Cyber Security at the La Trobe Law School, intercultural liaison for the Victorian Education Department's independent inquiry into antisemitism, co-chair of the Online Antisemitism working group of the Global Forum to Combat Antisemitism, an expert member of the Inter-Parliamentary Coalition to Combatting Antisemitism and served for two terms with the board of the UK's higher education regulator the QAA. His research interests include online regulation, hate speech and extremism in social media, and the impacts of technology on society.

He holds a PhD in Computer Science from Lancaster University, and a B. Comp. Sci. (Hons) & LLM(Juris Doctor) from Monash University. He is a Senior Member of the IEEE, a Graduate Member of the Australian Institute of Company Directors and a Member of the Victorian Society of Computers & Law.

PROF. DAVID WISHART

Dr David Wishart is a Director of the Online Hate Prevention Institute and an Adjunct Professor with the Law School at La Trobe University. He recently completed a term as Acting Dean of the Law School. His expertise includes Competition Policy, Corporations Law, Constitutional Law, Electronic Evidence, the law as to citizenship, and issues relating to law and Indigenous peoples. His *Curriculum Vitae* includes more than 50 refereed articles and a number of self-authored books.

He holds a Bachelor of Commerce, LLB(Hons), and an LLM from Melbourne University, and a PhD from the Australian National University.

SIMON KATTERL

Simon Katterl is an analyst at the Online Hate Prevention Institute and a consultant working in areas of human rights, mental health and community development. His written work is focused on human rights and regulatory oversight, with a particular focus on systems that interface with marginalised groups.

He holds a Bachelors of Law (Hons) and International Relations, a Graduate Diploma in Psychology and is undertaking a Masters in Regulation and Governance.

EXECUTIVE SUMMARY

The proposed legislation would represent an important step forward, but there are weaknesses in both its approach and implementation processes. Were these addressed, a safer online environment for all Australians would be secured.

The following are the key recommendations discussed in this submission:

Recommendation 1: As civil society can operate in a complementary fashion to government in delivering online safety, it is essential the legislation requires eSafety engage broadly in cooperation and consultation with, and support to, civil society organisations working in the online safety space.

Recommendation 2: The work of civil society organisations can be hampered by existing legislation as well as by certain private contract provisions in online platforms' terms of service. Introducing exemptions for civil society organisations and voiding certain contractual terms may be in the public interest.

Recommendation 3: Funding to support charities working towards online safety is needed. Such funding should be available broadly and not only in support of existing schemes under the Act.

Recommendation 4: It is essential terrorist materials, like manifestos and videos, also receive an "RC" rating to ensure they are not spread in Australia through other offline formats.

Recommendation 5: The ability to declare content illegal in Australia would itself be useful, as ancillary services will often ask that this be demonstrated before they will consider taking voluntary action.

Recommendation 6: Significant online harm also results from hate and incitement to violence that targets segments of the community, as distinct from the cyberbullying of individuals. These is a significant gap of coverage in this area.

Recommendation 7: We need to grant eSafety the power, working with the Australian Human Rights Commission, to counter and take down material which incites hate.

Recommendation 8: A takedown power covering incitement to hate, against both individuals and groups, is urgently needed.

Recommendation 9: We recommend joining the *Additional Protocol to the Convention on Cybercrime* which would allow greater cooperation with a range of other countries in tackling this problem

Recommendation 10: We recommend being more specific and stating the grant power under S 27(1)(g) is to be used: "to make, on behalf of the Commonwealth, grants of financial assistance in relation to online safety for Australians *including through education, training, online monitoring and reporting, efforts to enhance online safety, and the prevention and mitigation of online harms*;"

Recommendation 11: We support the inclusion of the civil penalty provision in the new scheme for cyberabuse of an adult, but strongly recommend the old scheme covering cyberbullying of a child be updated to include an equivalent penalty.

Recommendation 12: We recommend changing the wording of the image based abuse scheme so that it applies not just to the posting of an "intimate image" or "non consensual intimate image" but also an image **purporting** to be such an image of a particular person.

Recommendation 13: The new Bill unnecessarily broadens the scope of eSafety to refuse to investigate matters. By introducing clause 43(1) and the catch-all 43(2), the new legislative framework allows too much discretion with little oversight or accountability for decisions not to investigate.

Recommendation 14: The Abhorrent Violent Content scheme from the Criminal Code should be reviewed and exemptions updated when it is imported into the Online Safety Act.

Recommendation 15: Broadly, the provisions reflect higher thresholds for unlawful content – and therefore lower levels of protection – than existing Commonwealth, State and Territory protections against hate speech and vilification. This means we are committing to a lower standard of protection in our online life than our offline life. This should be reviewed.

Recommendation 16: The overly high barriers to use are resolvable by removing the intention element from both cyber-bullying and cyber-abuse provisions, so that this Act better aligns with agreed community standards in other Commonwealth, State and Territory provisions.

Recommendation 17: We recommend defining the ordinary reasonable person as being in the position of the person receiving the communication, or in instances where it targets someone based on an attribute such as race, gender, sexual orientation and religion, an ordinary reasonable representative of that group the attribute reflects.

Recommendation 18: We recommend a lower test in the cyber-bullying and cyber-abuse schemes would remove the term "serious", or would clarify that the meaning of serious harm was grounded in the experience of the individual or group that they represent.

Recommendation 19: We must recognise that the Internet is often a broadcast medium and protections are needed against hate and incitement that targets groups, not just individuals.

Recommendation 20: We recommend that the eSafety be granted referral powers to act on matters referred to it by other entities.

Recommendation 21: We need to go beyond core expectations supporting government regulation to include core expectations that will lead to the design of safer software. Online platforms should hear concerns from the Australian public, and taking the lead in improving online safety as broadly understood. The point in s 46(1)(f) about enabling reports for breaches of the terms of use of a particular platform does not address this gap as it gives no guidance as to what those terms of service

should cover, and this can create a perverse incentive to reduce protections in the terms of service and design systems to meet the minimum standards required by law, ultimately creating a less safe internet.

Recommendation 22: Greater oversight of the exercise of discretion and the use (or non-use) of powers by eSafety is necessary.

THE ROLE OF CIVIL SOCIETY

We welcome the Government's approach to online safety reform. We agree that industry must take responsibility for ensuring the digital services they offer in Australia are safe for Australians to use. We agree that legislation should reflect the primary responsibility of industry, and allow the regulator to assess this effort and act where industry falls short. We also agree that when Australians encounter online harms, it is essential they are sufficiently supported.

We agree the regulator should be resourced to provide online safety support but we believe there are limits to a regulator's capacities and that government cannot sufficiently resource a regulator to fully meet the community's need. Civil society has access and influence, both within the community and with the technology companies. These capacities are different from those of government and can operate in a complementary fashion. Accordingly, civil society has an essential role to play but, as in many other areas, its work must be partially resourced by government. This is already occurring in some areas of eSafety, but not in others. **It is essential the legislation requires eSafety engage broadly in cooperation and consultation with, and support to, civil society organisations working in the online safety space.**

The Online Hate Prevention Institute plays a unique role and has significantly enhanced the online safety of Australians, yet for every threat we address, the lack of government support for core operational capacity means other threats go unanswered.

It is through civil society that we learn of the changing expectations of the Australian community. It is through civil society that the community receives support in emerging areas that the government, including eSafety, have not yet addressed. Civil society, properly funded and given flexibility to meet the broad goals of increasing online safety for Australians, could provide a necessary backstop in this area of rapidly changing technology, harms, and community expectations on online safety.

The work of civil society organisations can be hampered by existing legislation as well as by certain private contract provisions in online platforms' terms of service. Introducing exemptions for civil society organisations and voiding certain contractual terms may be in the public interest. Such protections for civil society organisations could be limited to those organisations the government admits to a civil society online safety register. Alternatively, protections could be granted to charities on the existing Register of Harm Prevention Charities, many of which deal with harmful online content of one form or another. Even in area that might appear remote to online safety, such as harm prevention related to preventing and treating cancer, the promotion online of false information has put lives at risk.

Funding to support charities working towards online safety is needed. Such funding should be available broadly and not only in support of existing schemes under the Act. This will help ensure emerging issues, those not well covered by the proposed Online Safety Act, can still be addressed.

EFFICACY OF THE PROPOSED ONLINE SAFETY SCHEMES

The functions of eSafety, both as they currently stand and as proposed, are in our view too narrow to achieve its high-level objectives in a consistent manner. We are particularly concerned that terrorist material may not be sufficiently addressed. While that the draft Act provides less cumbersome approach to take down such material from the Internet is welcome, **it is essential terrorist materials, like manifestos and videos, also receive an "RC" rating to ensure they are not spread in Australia through other offline formats**. With respect to internet distribution, we welcome the Ancillary Service Scheme, although we believe it could be more strongly implemented. **The ability to declare content illegal in Australia would itself be useful, as ancillary services will often ask that this be demonstrated before they will consider taking voluntary action.**

Significant online harm also results from hate and incitement to violence that targets segments of the community, as distinct from the cyberbullying of individuals. These is a significant gap of coverage in this area. Attributes such as race, religion, mental or physical illness or disability, sexual orientation, gender identity/expression, intersex status and others are used to target segments of the community. In the most serious cases online hate against these groups involves incitement not only to hate, but also to violence. Preventing online harms requires action well before it reaches that point.

As Chancellor Angela Merkel stated after recent deadly attack in Frankfurt, Germany, , "Racism is a poison. Hatred is a poison. This hatred exists in our society and it is responsible for far too many crimes". **We need to grant eSafety the power, working with the Australian Human Rights Commission, to counter and take down material which incites hate.** Waiting until there are explicit threats of violence, or video from attackers who have documented their deadly violence, is too late. If that is the best we can do, we will have failed to protect Australians online.

According to recent research by the Australia Institute, around 12% of 18-24 year-olds and around 13% of 25-34 year-olds experienced online harassment in the form of "abusive language about your religious or ethnic background" while other age groups experienced the problem at slightly lower levels.[1] As a result of this particular type of abuse, 10% of those impacted lost more than 35 hours of work, 7% lost 14-35 hours, 4% lost 7-14 hours and 5% lost under 7 hours of work.[2] The actual figure within impacted communities will be far higher, as can be seen from Monash University research which found that 80% of the Australian Jewish community had seen online antisemitism in the preceding year, and while

[1] "Trolls and polls –the economic costs of online harassment and cyberhate", The Australia Institute, January 2019, p. 15 <https://www.tai.org.au/sites/default/files/P530%20Trolls%20and%20polls%20-%20surveying%20economic%20costs%20of%20cyberhate%20%255bWEB%255d_0.pdf>

[2] Ibid p. 16.

antisemitism was seen in a variety of forums, it was seen on Facebook specifically by 80% of 18-29 year-olds.[3] This is a significant form of harm impacting people's health and wellbeing, the public good of an inclusive society, as well as negatively impacting the Australian economy.

Other countries have been taking action on this problem as can be seen in the Netzwerkdurchsetzungsgesetz (Network Enforcement Act) in Germany, the European Commission Code of Conduct on Countering Illegal Hate Speech Online, the Canadian Digital Charter, and the French legislation on 'overtly hateful' content. We believe **a takedown power covering incitement to hate, against both individuals and groups, is urgently needed.** This would go beyond the current proposal of a cyberbullying provisions for adults and would remove the idea that an individual must be identified. It would also go beyond the incitement to violence provision. In implementing this system, there is a need for an interface between the Australian Human Rights Commission and eSafety.

We note the proposed transparency reporting will include data related to platforms' responses to online hate. Australia, like other countries should have powers to have such hate speech taken down. **We recommend joining the *Additional Protocol to the Convention on Cybercrime* which would allow greater cooperation with a range of other countries in tackling this problem.**[4]

The grant making function under S 27(1)(g) and S 27(2) has not changed and covers "assistance in relation to online safety for Australians". In practice the use of this power has been limited to [5] **We recommend being more specific and stating the grant power under S 27(1)(g) is to be used: "to make, on behalf of the Commonwealth, grants of financial assistance in relation to online safety for Australians *including through education, training, online monitoring and reporting, efforts to enhance online safety, and the prevention and mitigation of online harms*;"** This would see the grant promote both objectives of the Act (S 3) where at present only the "promotions of online safety" part is being addressed. Enabling Civil Society to take greater action and to work in partnership with eSafety is the most impactful and cost-effective way to increase online safety.

COMMENTS ON THE PROPOSED SCHEMES

DIFFERENCES BETWEEN CYBER BULLYING SCHEMES

There are differences between that do not have a strong basis between the schemes for cyberbullying of a child and for cyberabuse of an adult. One difference is in Part 7, S 89, of the Bill which gives the Commissioner power to issue an end user with a removal notice in relation to cyberabuse of an adult,

[3] David Graham & Andrew Markus, *Gen17 Australian Jewish Community Survey: Preliminary Findings*, 2018, p. 60—70 <https://www.monash.edu/__data/assets/pdf_file/0009/1531791/gen17-initial-findings-report-online-version-final-22_3.pdf>

[4] *Additional Protocol to the Convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, opened for signature 28 January 2003 (entered into force 1 March 2006)

[5] https://www.communications.gov.au/what-we-do/internet/online-safety/support-online-safety

which if not complied with can result in a civil penalty under S 91. The equivalent provision in relation to end users in relation to cyberbullying material targeting a child is in under Part 5, S 70, of the Bill with if not complied with triggers S 71, only S 71 has no civil penalty provision and the only remedy is via a court order for an injunction under S 165, which in fact also applied to a breach of S 91. This means the Commission has *less* power in relation to material targeting children than in relation to material targeting adults. **We support the inclusion of the civil penalty provision in the new scheme for cyberabuse of an adult, but strongly recommend the old scheme covering cyberbullying of a child be updated to include an equivalent penalty.**

The bullying of children by adults, and of adults by children, is well documented. If the intent is to protect children from penalties from failing to comply with an enforceable removal notice, it would be far better to make notices enforceable for cyberbullying of both adults and children, and then to exempt children from enforcement action for failing to comply or to impose a different penalty (though S 165 may already allow for this via an injunction). As it stands, if a 17 year old and an 18 year old engage in cyberbullying against each other, it is the 17 year old who can be subject to an enforceable removal notice regarding the content they posted, while the 18 year old, because they are bullying a child, are not subject to an forceable removal notice.

IMAGE BASED ABUSE SCHEME

We recommend changing the wording of the image based abuse scheme so that it applies not just to the posting of an "intimate image" or "non consensual intimate image" but also an image purporting to be such an image of a particular person. The harm can be very similar even if the image is not in reality of the stated person. This may be a rarer occurrence, but the law would be more complete if it were able to take account of this.

Having the same penalty for an individual and a company is unusual and likely to be either disproportionately high for individuals, or so low as to simply be absorbed by companies. Typically, corporate penalties are five times higher than those for individuals, which makes sense given the vastly different level of resources at stake. A penalty expressed as the higher of an absolute value or a percent of revenue for a company would be best. Alternatively, different penalty rates for individuals and corporations, with the corporate penalty being 5 or more times higher, would be more appropriate.

ONLINE CONTENT SCHEME

The new Bill unnecessarily broadens the scope of eSafety to refuse to investigate matters. By introducing clause 43(1) and the catch-all 43(2), the new legislative framework allows too much discretion with little oversight or accountability for decisions not to investigate.

In the past we have struggled to get action from eSafety to exercise their existing powers to address content that breaches the online content scheme. For example, on one occasion we contacted eSafety to ask for a refer of a Terrorist manifesto to the Classification Board. This was a very similar circumstance to the Christchurch manifesto which was referred (it involved a live streamed deadly attack), but eSafety did not take action and as a result the content was not given an "RC" classification.

In fact, last we checked the Christchurch manifesto was the *only* terrorist manifesto referred to the Classification Board.

The failure to take action led to advertising for a government program appearing alongside the terrorist manifesto.[6] Hence taxpayers' money provided profits for an individual hosting material which incited deadly violence and which had been produced as part of a deadly terrorist attack overseas. We need stronger action and less discretion when it comes to such clear-cut material putting Australian lives at risk.

Decisions to not exercise powers and fulfill functions by eSafety undermine the objectives of legislation and reduce online safety for the community. Therefore, the failure to exercise of powers should be given greater clarity and accountability, rather than greater discretion. Clause 43 should be removed from the Bill and replaced with an obligation on eSafety to provide written reasons to complainants and public disclosures in their annual reports, on decisions not to investigate complaints brought to them under this scheme.

We note the reading guide does not make mention of powers in relation to search engines or other linking services, but that this remains covered in the legislation. We welcome this continued coverage. Added clarity that it applies regardless of where a search engine is based or content is hosted would be welcomed. It was suggested to us at one point by eSafety that Google was exempt from the legislation, a point we disagree with on technical grounds given the known presence of certain equipment in Australia, but which under eSafety's interpretation would fully largely defeat the legislation's purpose given the size of Google's market share. The terrorist manifesto we found hosted along-side government advertising was listed in Google.

ABHORRENT VIOLENT CONTENT SCHEME

This Abhorrent Violent Content scheme mirrors provisions in the Criminal Code, however those provisions were created in a rush and are flawed.[7] For example, the journalism protection in s 104(1)(e) will not cover community news service run by volunteers, nor will it cover interns at professional media organisations. For eSafety purposes there is a valid question over whether it should cover at least some blogs or podcasts as online news services. In the online space the *professional journalist* in many ways has less meaning. The news and opinion articles by academics via *The Conversation* is a case in point.

An explicit exemption for civil society organisations working in the space of countering violent extremism would be welcome. The exemption for "conducting research" may or may not cover publishing the results of research. There is no exemption for the use of such material for teaching

[6] https://ohpi.org.au/advertising-supports-hosting-of-terrorist-manifesto/

[7] https://theconversation.com/new-livestreaming-legislation-fails-to-take-into-account-how-the-internet-actually-works-114911

purposes, for example publishing it on an online learning management system restricted to students studying e.g. counter terrorism.

The Abhorrent Violent Content scheme from the Criminal Code should be reviewed and exemptions updated when it is imported into the Online Safety Act.

MEETING COMMUNITY STANDARDS

Addressing cyber-bullying and cyber-abuse are commendable intentions of this Bill. They should however, be understood within the context of broader hate speech and vilification that disproportionately target particular communities .

The legal thresholds for lawful and unlawful content should reflect an understanding of and commitment to addressing this problem. Unfortunately, the thresholds outlined for cyber-bullying (clauses 5 and 6) and cyber-abuse (clauses 5 and 7), do not achieve this.

Broadly, the provisions reflect higher thresholds for unlawful content – and therefore lower levels of protection – than existing Commonwealth, State and Territory protections against hate speech and vilification. This means we are committing to a lower standard of protection in our online life than our offline life. This should be reviewed.

There are three key issues that make the cyber-bullying and cyber-abuse thresholds too high to provide adequate protection to Australians. First, they require an intention element – meaning the person *intended* to cause harm – which is not required in other protections. Second, the test fails to clearly articulate from what vantage point the "reasonable person" stands in assessing whether something is harmful – the person making the statement, or the person receiving it? Third, the level of harm required is simply far too high, given online platforms often require evidence material is in breach of a specific law before they will take action, this leaves Australians, including our Australian children, unprotected.

ISSUE 1: INTENTION ELEMENT

Both cyber-bullying and cyber-abuse provisions make intention an element of unlawful content. This is an unreasonable and out-of-step approach on two grounds.

First, this goes beyond existing Commonwealth, State and Territory provisions – both civil and some criminal – which do not require an intention element.[8] For example, existing Commonwealth Criminal Code provisions (*Criminal Code Act 1995* (Cth), art 474.17) that make it an offence to menace and harass

[8] For civil examples with lower thresholds, see: *Racial and Religious Tolerance Act 2001* (Vic) ss 7,8; *Anti-Discrimination Act 1977* (NSW) ss20C, 38S, 49ZT and 49ZXB; *Anti-Discrimination Act 1991* (Qld) s 124A(1); *Anti Discrimination Act 1998* (Tas) ss 17(1) & 19; *Civil Liability Act* 1936 (SA) s 73(2); *Equal Opportunity Act 1984* (WA) ss49A-49C. While some criminal provisions such as ss 24 and 25 of the *Racial and Religious Tolerance Act 2001* (Vic) appear to have an intention element, others such as s 474.17 of the *Criminal Code Act 1995* (Cth) makes certain communications an offence without an intention element from the communicator, and is instead focused on the harm caused to the recipient.

a person do not require an intention element, and have a penalty of up to three years imprisonment. So, it is easier to send someone to prison for 3 years under existing laws than it is to remove content under the proposed laws. This clearly does not meet the objectives of the Bill.

Second, it unfairly limits the protections by allowing communicators of hate speech and vilification to continue to do so, with either ignorant or bad-faith claims that they intended no harm. This latter issue it particularly important for marginalised communities, such as the LGBTIQA+ or Aboriginal and Torres Strait Islander communities who receive harmful messages with either misguided benevolent intentions, or with little regard or consideration for their experience of those messages.

The overly high barriers to use are resolvable by removing the intention element from both cyber-bullying and cyber-abuse provisions, so that this Act better aligns with agreed community standards in other Commonwealth, State and Territory provisions.

ISSUE 2: REASONABLE PERSON TEST

The "ordinary reasonable person" test articulated in both clauses and further clarified in clause 8 states that in determining whether content is offensive – and potentially constituting cyber-bullying or cyber-abuse – particular regard should be given to the "standards of morality, decency and propriety generally accepted by reasonable adults", as well as two later criteria regarding the literary, artistic, or educational merit of the content, and its general character (e.g. medical, legal or scientific).

This test, while on its face helpful, does not grapple with the divergence in our community between people who often communicate hate speech and vilification, and those who fall victim to it. What idea of the ordinary reasonable person encompasses both these groups – for example in instances of discredited Holocaust denial groups and Jewish Australians – remains unclear. This will leave a great deal of uncertainty for all groups about what forms of civic and online engagement are appropriate. In this way, the "objective test" in clause 8 will devolve into a confusion of subjective assessments from different groups about what is reasonably considered to be offensive or immoral.

We recommend defining the ordinary reasonable person as being in the position of the person receiving the communication, or in instances where it targets someone based on an attribute such as race, gender, sexual orientation and religion, an ordinary reasonable representative of that group the attribute reflects. Such an approach would take into account the values standards and other circumstances of that group – such as historical experiences of genocide and colonialism, along with contemporary experiences of discrimination – which are the critical factors that determine offence and cause harm.

ISSUE 3: HIGH HARM THRESHOLDS, LOW PROTECTION

The cyber-bullying (cl 6) and cyber-abuse (cl 7) provisions place high thresholds in order to determine whether content is unlawful. Clause 6(1)(b)(ii) provides that, in addition to an intention element, the content would:

be likely to have the effect on the Australian child of seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child.

Meanwhile, clause 7(1)(b) – in addition to the "offense" test noted above in clause 7(1)(c) – requires that the content was intended to cause *serious harm*, which is somewhat tautologically defined in clause 5 as causing serious harm to a person's physical or mental health (temporary or permanent).

In both cyber-bullying and cyber-abuse tests, the threshold is far too high. Our children may be subject to harmful content that: 1) is prolonged and compounding, leading to isolation, but not meeting the acute test of "seriously" intimidating, harassing or humiliating; and, importantly, 2) the communicators of that content – in particular children – might not do so with the intention of causing serious mental or physical harm, but nevertheless do cause it.

A simpler approach, in addition to removing the intention element and amending the ordinary reasonable person element, is to lower these thresholds. This is particularly critical for children. **We recommend a lower test in the cyber-bullying and cyber-abuse schemes would remove the term "serious", or would clarify that the meaning of serious harm was grounded in the experience of the individual or group that they represent.**

PROPOSED TESTS

We propose that content eligible under 6(1)(a) would be cyber-bullying if:

(b) an ordinary reasonable person – in the position of the person receiving material – would conclude that:

(i) the material is threatening, intimidating, harassing or humiliating the Australian child;

(ii) …

(c) such other conditions (if any) as are set out in the legislative rules;

Likewise, content that was eligible under clause 7(1)(a) would be deemed **cyber-abuse** (regarding adult protections) if:

(b) an ordinary reasonable person – in the position of the person receiving the material – would identify the material as seriously harmful, menacing, harassing or offensive.

(c) such other conditions (if any) as are set out in the legislative rules…

A similar amendment would be made to clause 8(1) so that it stated "the matters to be taken into account in deciding for the purposes of this Act whether an ordinary reasonable person – **in the position of the person receiving the material –** would regard the material as being, in all the circumstances, offensive (including the existing sub-provisions (a) through to (c)).

PROPOSAL FOR INCLUSION OF REFERRAL POWERS

The above approach still leaves significant areas of online hate speech and vilification unprotected. There is a range of State and Commonwealth hate speech and vilification provisions that protect specific communities. This is important for giving visibility and protection to those communities and individuals from those communities. Importantly, most of these legislative protections have lower thresholds to deem content hate speech or vilification, and therefore are likely to provide greater protection to community members.

The following is a table that sets out such legislation and indicates whether the proposed Online Safety Act provides the same or further protection when an individual is personally attacked online on the basis of a protected attribute of their identity:

Legislation	**Who is protected?**	**Is the standard of protection in the proposed Act higher, similar or lower?**
Racial and Religious Tolerance Act 2001 (Vic), ss 7 and 8	People who experience vilification based on race and religion.	**Similar standard of protection**
Anti-Discrimination Act 1977 (NSW) ss 20C, 38S, 49ZT and 49ZXB	People who experience vilification based on race, transgender status, homosexuality, and HIV/AIDS status.	**Similar standard of protection**
Anti-Discrimination Act 1991 (Qld) s 124A(1)	People who experience vilification based on race, religion, sexuality or gender identity.	**Similar standard of protection**
Discrimination Act 1991 (ACT) s67A(1)	People who experience vilification based on disability, gender identity, HIV/AIDS status, race, religious conviction, sex characteristics, and sexuality.	**Similar standard of protection**
Anti-Discrimination Act 1998 (Tas) s 17	People who experience certain conduct (hate speech) based on their age, race, disability, sexual orientation, lawful sexual activity, gender, gender identity, intersex variations of sex characteristics, pregnancy, breastfeeding, marital status, relationship status, family responsibilities and parental status.	**Lower standard of protection**
Racial Vilification Act 1996 (SA) s 4	People who experience racial vilification.	**Similar standard of protection**
Equal Opportunity Act 1984 (WA) ss 49A-49C	People who experience racial harassment in employment, education and accommodation.	**Lower standard of protection**

Racial Discrimination Act 1975 (Cth) s 18C	People who experience hate speech based on race.	**Lower standard of protection**
Criminal Code Act 1995 (Cth) s 474.17	People who are menaced, harassed or offended by someone using a "carriage service"	**Lower standard of protection**

Significantly, however, most of the above protections also cover harm targeting "a group" of people defined by the protected attribute. The provisions in this act lacks such coverage. Material imputing that all Holocaust survivors are liars would be actionable under most other acts, but not under this act unless a specific Holocaust survivor was targeted. **We must recognise that the Internet is often a broadcast medium and protections are needed against hate and incitement that targets groups, not just individuals.**

In order to address such lack of coverage, **we recommend that the eSafety be granted referral powers to act on matters referred to it by other entities.** These entities should include civil society organisations acting to prevent online harms, and State and Commonwealth government departments and agencies, such as human rights and equal opportunity commissions, police and other emergency services.

Referral powers would enable the Safety to:

- leverage investigative material that is collected and compiled by other entities, and provided through the referral mechanism;
- ensure that take-down notices are issued to the correct body, and executed in a consistent and thorough manner;
- issue takedown notices for online material which passes thresholds of tolerance for other state or federal legislation, such as racial vilification, counter-terrorism or criminal legislation; and
- participate in coordinated campaigns with other entities to exercise eSafety's conferred function of "Protecting Australians from Terrorist or Violent Criminal Material."

Referral powers also provide an interface for collaboration, which would allow the eSafety and the other entities to concentrate their operations within their specific areas of expertise, thus avoiding duplication of effort and ensuring the more efficient use of resources to defend against the rising threat to the online safety of all Australians.

PROPOSAL FOR SAFETY BY DESIGN COMPLIANCE

The core expectations in section 46 help to ensure compliance by design by requiring mechanisms be built in to facilitate, for example, reporting. We welcome this, but feel ss 46(1)(c) and (e) are substantially incomplete. As these standards are not legally binding (see s 45) is it disappointing that common societal expectations, which most platforms already include in their community standards, are not included here.

Not making such standards in compliance by design sends a message that opposition to racism, misogyny, homophobia, the promotion of Nazism, white supremacy or other extremist ideologies, is **not**

considered a core expectation of online safety by the Australian government. This sets a lower standard than industry sets itself. Safety by design would be significantly advanced by stating the expectations in these and other areas. There are existing expectations as a result of legislation at both Commonwealth and State level, and decisions by the courts. The standards and tests do not need to be outlined in this legislation, but reference to the fact that Australia expects the reporting mechanisms available to users be capable of accepting reports of, e.g. racist content, is an important requirement in the context of promoting safety by design. Similarly, the bar for taking legal action on extremist content is set unnecessarily high for a reporting mechanism.

We need to go beyond core expectations supporting government regulation to include core expectations that will lead to the design of safer software. Online platforms should hear concerns from the Australian public, and taking the lead in improving online safety as broadly understood. The point in s 46(1)(f) about enabling reports for breaches of the terms of use of a particular platform does not address this gap as it gives no guidance as to what those terms of service should cover, and this can create a perverse incentive to reduce protections in the terms of service and design systems to meet the minimum standards required by law, ultimately creating a less safe internet.

PROPOSAL FOR INCLUSION OF SPECIFIC OVERSIGHT

Without enforcement, legislation is unlikely to protect individuals and communities. As discussed, we have had prior experiences that indicate eSafety's reluctance to take up investigations or pursue content removal requests. These may be a small window into a larger issue about the exercise of discretion contrary to the spirit of protecting our community, potentially due to significant under resourcing.

Greater oversight of the exercise of discretion and the use (or non-use) of powers by eSafety is necessary. There are several ways that this could be achieved, including:

- An obligation on eSafety to provide written reasons to complainants about why their matter was not investigated
- Annual obligations on eSafety to report on decisions regarding the exercise or restraint from exercising powers to investigate under the Act. This should include transparency reporting including details of how many items were removed under each provision of the Act. Where people are targeted based on a specific characteristic, such as race, religion, sexuality, disability, etc., the report transparency report should provide disaggregated information on the number of reports and removals for each attribute. Further details could be provided through Ministerial Directions.
- An external review into eSafety within 3 years of the Act (and every 3 years following), examining whether eSafety is achieving the objectives of the Act, and whether it is appropriately balancing different rights and obligations as set out under the Act. Such approaches to oversight should include civil society, and invite feedback from the public, with transparent reporting of the results of that feedback.