LA LSIC - AVP INQUIRY
SUBMISSION NO. 16
RECEIVED 19 DECEMBER 2019

Anti-Vilification Protections Inquiry

Office of the eSafety Commissioner submission

19 December 2019

Acknowledgement

eSafety acknowledges the Aboriginal and Torres Strait Islander peoples as the Traditional Custodians of this country and recognise their ongoing connection to land, wind, water and community. We pay our respect to Aboriginal and Torres Strait Islander peoples, their cultures and to Elders past, present and emerging.

Introduction

When the *Racial and Religious Tolerance Act 2001* (Vic) (Racial and Religious Tolerance Act) was introduced in 2011, Facebook was still three years away from being founded and the first iPhone was still six years away from being released.

Since that time, internet and digital technologies have revolutionised the lives of Australians. While this undeniably brings enormous benefits and opportunities, it also presents risks and harms, including cyber abuse and online vilification.

Both cyber abuse and online vilification are fundamentally manifestations of social, cultural and behavioural issues that defy single or simple solutions.

Yet as Australia's leader in online safety, the Office of the eSafety Commissioner (eSafety) knows that regulation also has a crucial role to play in shaping behaviour and providing a safety net and means of redress for Australians who experience harm online.

eSafety welcomes the opportunity to provide a submission to the Victorian Legal and Social Affairs Committee's (Committee) Inquiry into Anti-Vilification Protections (Inquiry).

Terms of Reference

eSafety notes that the Racial and Religious Tolerance Act applies to conduct both online and offline. eSafety's submission focuses on the Terms of Reference relating to online safety and vilification laws as they apply to online vilification.

eSafety's submission provides insights from its regulatory experience and in-depth research and evidence base. eSafety hopes these insights, especially into the intersectional nature of cyber abuse, will both demonstrate and support its two key recommendations:

1. Expand the grounds for vilification, including online vilification, under the *Racial and Religious Tolerance Act 2001* (Vic) to account for the intersectional nature of online abuse and harms.

2. Recognise and promote the services of eSafety's services and programs, as part of a multi-faceted and holistic approach to addressing online harm, including online vilification.

Regulatory approach

eSafety is the only government agency in the world dedicated specifically to online safety. It leads, coordinates, educates and advises on online safety issues to ensure all Australians have safe, positive and empowering experiences online.

eSafety adopts a whole of community and multifaceted regulatory approach, which draws upon social, cultural, technological and regulatory initiatives and interventions.

In September 2019, eSafety released its 2019-2022 Strategic Plan. This outlined eSafety's Mission, Vision and Values. It also highlighted that promoting diversity, inclusion and resilience online is one of eSafety's key priorities.

eSafety's regulatory approach recognises the intersectional nature of cyber abuse.

At its core, intersectionality explores the complex and cumulative intersection of various forms of social discrimination and inequality. It explores how socially constructed and maintained forms of oppression can create discrimination and inequality on both a systemic and interpersonal level. Ultimately, intersectionality recognises the multi-dimensional and contextual nature of an individual's experience of discrimination and inequality, while also providing a lens for examining the dynamics of power and oppression across different situations and settings in society.

It is important not to portray those who experience intersectional discrimination and inequality as inherently vulnerable or marginalised. Rather, it is the social modes of oppression, such as sexism, racism, ableism, ageism, homophobia and transphobia, which are risk factors that may increase an individual's risk of experiencing harm online.

eSafety adopts an inclusive and strengths-based approach to diversity. This recognises that diversity and identity can be protective factors that promote resilience and can be leveraged to help people better navigate digital environments. It also recognises that online safety responses cannot be homogeneously applied across social and cultural groups. Rather, they must be tailored, inclusive and developed through co-design.

For example, eSafety recently developed a new online resource to assist Elders and Aboriginal and Torres Strait Islander women help women within their communities who are experiencing technology-facilitated abuse. This was developed based on eSafety research that showed Aboriginal and Torres Strait Islander women are experiencing online abuse, restricted access to technology and stalking as part of a wider pattern of domestic and family violence. However, it was also developed based on the strengths of Aboriginal and Torres Strait Islander cultural practices of family life and kinship, including the pivotal role of Elders within Aboriginal and Torres Strait Islander communities.

Ultimately, by recognising the multiple factors that intersect to shape an individual's experience online, eSafety can recognise risk factors that may shape an individual's risk of cyber abuse and leverage protective factors that are likely to assists individuals navigate online harms and build resilience, while promoting diversity and inclusion online and in society.

Interaction between state, territory and Commonwealth legislation

Laws relating to online harms, including online vilification, exist at the Commonwealth, state and territory levels.

At the Commonwealth level, cyberbullying under the *Enhancing Online Safety Act 2015* (Enhancing Online Safety Act) is online behaviour that an ordinary reasonable person would conclude is likely to have a seriously threatening, seriously intimidating, seriously harassing or seriously humiliating effect on an Australian child.[1]

eSafety has also adopted this definition for cyber abuse, an umbrella term to capture different types of online abuse that apply more broadly to include adults.[2] For example, serious harassment online may constitute cyber abuse.

At the Victorian state level, vilification applies to two grounds under the Racial and Religious Tolerance Act: race and religion.[3] Further, there are two categories of vilification offences: unlawful vilification and serious vilification.[4] In summary, unlawful vilification occurs when a person, on the grounds of either race or religion, engages in conduct that incites hatred against, serious contempt for, or revulsion or severe ridicule of, that other person or class of person.[5] Broadly speaking, serious vilification occurs when a person, on the grounds of either race of religion, intentionally engages in conduct that the person knows is likely to incite hatred, threaten physical harm or incite serious contempt for, or revulsion or severe ridicule of, that other person or class of persons.[6] The key difference between the two categories of vilification offences is the element of intent.

eSafety believes its reporting schemes under the Enhancing Online Safety Act and Victoria's anti-vilification laws work in a complementary and mutually reinforcing manner.

[1] *Enhancing Online Safety Act 2015* (Cth) section 5(1)(b)
[2] While eSafety does not have formal powers under the *Enhancing Online Safety Act 2015* (Cth) relating to adult cyber abuse, depending on the nature, context and severity of the online abuse, eSafety may be able to draw upon its cooperative arrangements with social media services to informally provide relief to the individual.
[3] *Racial and Religious Tolerance Act 2001* (Vic) section 1
[4] *Racial and Religious Tolerance Act 2001* (Vic) Parts 2 and 4
[5] *Racial and Religious Tolerance Act 2001* (Vic) sections 7 and 8
[6] *Racial and Religious Tolerance Act 2001* (Vic) sections 24 and 25

This is demonstrated by highlighting several of the key differences between cyber abuse and online vilification. Notably, the different thresholds give rise to different, but ultimately complementary, regulatory responses.

First, the threshold for online vilification is higher than cyber abuse. As such, any behaviour that meets the bar for online vilification would constitute cyber abuse under eSafety's framework. However, given the appropriately higher bar to constitute vilification, not all cyber abuse would constitute online vilification. Whether cyber abuse constitutes online vilification will therefore depend on the nature, context and severity of the cyber abuse.

Second, cyber abuse is not limited to the grounds of race and religion. eSafety's scheme does not require that the online material falls within a particular ground, but rather that it satisfies the definition outlined above. eSafety could therefore handle cyber abuse complaints that relate to grounds that do not fall under the Racial and Religious Tolerance Act, such as sex, sexual orientation, gender identity and disability.

The distinction in thresholds between the regulatory schemes allows anti-vilification laws to be tightly focused on behaviour of the highest harm and eSafety's operations to have a broader focus on a wider range of online harms.

However, there is a discrepancy in that the Racial and Religious Tolerance Act is limited to particular grounds, which as outlined below, do not necessarily correlate to the most common experiences of online harm. eSafety believes the grounds needs to be expanded to account for the intersectional nature of online abuse and harms.

Recommendation: Expand the grounds for vilification, including online vilification, under the *Racial and Religious Tolerance Act 2001* (Vic) to account for the intersectional nature of online abuse and harms.

eSafety's complaints data and research

eSafety undertakes an extensive research program to ensure its programs and resources are evidence based. This equips eSafety with the insights and knowledge it needs to understand the nature of online safety issues and design, implement and evaluate best possible solutions.[7]

[7] The full suite of eSafety' research findings, as well as all the research referenced throughout this submission, is available at eSafety's Research Library at https://www.esafety.gov.au/about-us/research

Online hate speech

Of particular relevance to the Committee is eSafety's research into online hate speech, which is to be released in January 2020.[8] This research is based on the 2019 National eSafety Adult Online Safety Survey and is one of eSafety's more explorative research pieces to date. Rather than being based around a legislative definition of cyber abuse, it explores how adult Australians understand hate speech. The research shows that people understand the term broadly. It extends beyond the incitement or spreading of hate speech to communication that is hurtful, offensive or negative.

To further explore this understanding within the research design, eSafety developed the following broad question relating to hate speech:

> In the last 12 months, how many times, if ever, have you received a digital communication that offended, discriminated, denigrated, abused and/or disparaged you because of your personal identity/beliefs (e.g. race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, etc)?

Based upon the responses to this question, it is estimated that around 1 in 7 adult Australians aged 18–65 (14%) were the target of online hate speech in the 12 months to August 2019. Staggeringly, this is around 2 million people. People identifying as LGBTQI or Aboriginal or Torres Strait Islander experience online hate speech at double the national average.

In general, people experiencing online hate speech identify their political views (21%), religion (20%) and gender (20%) as the top three reasons for being targeted. However, terms like race, ethnicity and nationality can be interchangeable from a respondent's perspective. Taken together, these reasons account for 32% of people experiencing online hate speech. The findings were particularly stark in relation to sexual orientation. Sixty-one per cent of those who identify as LGBTQI report that their sexual orientation was the reason for being the target of online hate speech, compared to their gender (35%) or political views (30%).

Most people were unable to attribute responsibility for their online hate speech experience to a specific person, with 47% assigning blame to a stranger and 13% reporting that they didn't know who is responsible (13%). People identifying as LGBTQI (76%) are considerably more likely to identify a stranger as the source of online hate speech than any other group.

An estimated 58% of those personally experiencing online hate speech report a negative impact from their experience. Thirty-seven per cent report mental or emotional stress as a result of their experience, while 10% report reputational damage. People identifying as

[8] Office of the eSafety Commissioner, Online Hate Speech Research, to be released in January 2020

LGBTQI were more likely to report mental or emotional stress from online hate speech than other groups.

In the 12 months to August 2019, an estimated five per cent of the adult population in Australia (just over 738,000 people) intentionally visited an online site or forum hosting material targeting people.

While this research is both troubling and confronting, it also shows Australians want change. Seven in ten adult Australians believe that online hate speech is spreading, with the majority agreeing that more should be done to stop its growth either through the introduction of new laws (71%) or through social media companies doing more (78%).

Gender and intersectionality

eSafety's recent online hate speech research builds upon a broader body of eSafety research and complaints data that highlights the intersectional nature of cyber abuse.

More than two-thirds of eSafety's complaints about cyber abuse and image-based abuse involve women.

The personal, sexual and gendered abuse women receive online is part of a broader spectrum of violence, abuse and harassment against women. Reflecting society's broader gender inequality, often women are targeted online because they are women.

Applying an intersectional lens shows how gender intersects with other forms of social discrimination and inequality to shape different online experiences for women.

eSafety's 2017 *Image-based abuse: National survey - summary* report indicates that 1 in 10 Australians aged 18 years and over have had their intimate image/s or video/s shared without their consent.[9] This increases to 1 in 4 women between the ages of 18–24 and 1 in 5 for those identifying as LGBTIQ. 1 in 4 Aboriginal and Torres Strait Islander people have experienced image-based abuse.

In 2019, eSafety released two research pieces focusing on the intersectional nature of cyber abuse relating to women.

In February 2019, eSafety released research into the experiences of technology-facilitated abuse among women from culturally and linguistically diverse (CALD) backgrounds.[10] It

[9] Office of the eSafety Commissioner, Image-based abuse: National Survey - Summary Report, October 2017
[10] Office of the eSafety Commissioner, eSafety for Women from Culturally and Linguistically Diverse Backgrounds, February 2019

showed that the impacts of technology-facilitated abuse on CALD women are not substantially different to the impacts experienced by non-CALD women. However, social isolation may be amplified for CALD women where fear of shaming is particularly strong. Further, it showed that in some cases, perpetrators used culturally-specific threats, including: threats of deportation (especially for women on spousal visas); threats of honour killing delivered via a third party culturally-specific humiliation (such as sending images of a woman without her hijab); and threats of withholding Islamic divorce.

In October 2019, eSafety released research relating to Aboriginal and Torres Strait Islander women's experiences of technology-facilitated abuse.[11] Recognising that there is considerable diversity in the living circumstances, service access and needs of Aboriginal and Torres Strait Islander women regardless of where they live, this research focused on Aboriginal and Torres Strait Islander women living in rural areas. It was based on qualitative research comprised of in-depth interviews and a discussion group with service providers who support Aboriginal and Torres Strait Islander women.

Similarly to the research with CALD women, Aboriginal and Torres Strait Islander women did not experience substantially different impacts of technology-facilitated abuse from those felt by other women. However, some impacts appeared to be amplified, including the risk of being socially isolated from kinship networks and fear of shaming and family retribution. There were also culturally-specific service barriers to seeking support for technology-facilitated abuse, including: justice system barriers; issues with police; challenges in providing sufficient evidence; fear of racial prejudice and fear of police brutality; barriers related to child protection and courts; and legislation issues.

Online harms

As noted above, eSafety's online hate speech research showed that close to 60% of people who experienced hate speech reported a negative impact from their experience, typically mental or emotional stress, relationship problems or reputational damage.

This is sadly consistent with a range of eSafety research that highlights the devasting harms and impacts of cyber abuse.

eSafety's research shows that the impact of image-based abuse amongst online adults was overwhelming negative.[12] Women were considerably more likely to report negative personal

---

[11] Office of the eSafety Commissioner, Online safety for Aboriginal and Torres Strait Islander women living in urban areas, October 2019
[12] Office of the eSafety Commissioner, Image-based abuse: National Survey - Summary Report, October 2017

impacts as a result of image-based abuse, including in terms of their emotional response, fear of discovery and impact on life.

Overall, two-thirds of online adults who experienced image-based abuse indicated they had felt annoyed (65%) or angry (64%) with the person who had perpetrated the abuse, while many also felt humiliated (55%) and depressed (40%). Thirty-two percent had felt afraid for their safety. Online adults most feared the discovery of the photos/videos by their friends (51%) and family (48%), although many also feared discovery by an employer (41%), intimate partner (40%) and children (39%).

The most common negative impacts of the most recent experience of image-based abuse related to self-esteem (42%) and mental health (41%). One-third said it had impacted their physical wellbeing (33%) and relationships with friends (33%), while just over one-quarter said it had impacted their intimate/sexual relationships (28%), relationships with family (27%) and performance at work or study (28%).

What this ultimately demonstrates is that while cyber abuse can vary in form, it is consistent in the devastating harm it cases. eSafety thus supports broad online vilification protections that are commensurate to the harms being suffered and not limited to only several of the intersectional grounds that eSafety's research show correlate with disproportionately higher experiences of cyber abuse.

eSafety's role in addressing online harms

While broad anti-vilification laws are important, they are only one component of a multi-faceted approach to reducing harm online. As Australia's leader in online safety, eSafety has an integral role to play in addressing a broader range of online harm than online vilification. This is particularly important given that online harm exists on a spectrum and addressing harms at the lower end of the spectrum can help ensure they don't facilitate, or escalate into, to harms at the more extreme end of the spectrum, such as online vilification.

Recommendation: Recognise and promote eSafety's services and programs as part of a multi-faceted and holistic approach to addressing online harm, including online vilification.

Reporting schemes

Since 2015, eSafety has administered a complaints service for Australian children who experience serious cyberbullying. The scheme serves as a safety net for young people who haven't been able to resolve their online issue via the social network's reporting functions.

eSafety works closely with social media services to help remove harmful material and provide relief for a young person and their family.

Since the introduction of the scheme, eSafety has received over 1700 complaints about cyberbullying affecting Australian children. This includes over 500 complaints received this year by mid-November this calendar year, which represents an approximately 49% increase from the same time in 2018.

In July 2017, eSafety's remit was expanded to cover all Australians. However, because the cyberbullying legislative scheme was not extended at that time to cover adults, if a person over 18 years is experiencing serious cyber abuse, but their circumstances do not fall within one of our complaints or reporting schemes, eSafety focuses on providing information, guidance and support. Depending on the nature, context and severity of the online abuse, eSafety may be able to draw upon its cooperative arrangements with social media services to informally provide relief to the individual. This may involve getting cyber abuse material removed. As at mid-November 2019, eSafety had already helped over 800 adults in 2019, which represents an approximate 48% increase from the same time in 2018.

Relevantly, on the 11 December 2019, the Minister for Communications, Cyber Safety and the Arts, the Hon Paul Fletcher MP, launched the Australian Government's consultation process on a new Online Safety Act. Of the many proposals, which will considerably expand eSafety's role and remit if implemented, it includes the proposal to introduce a new cyber abuse scheme for Australian adults to facilitate the removal of serious online abuse and harassment and introduce a new end user take-down and civil penalty regime.[13] eSafety strongly supports this proposal.

In October 2017, eSafety introduced an image-based abuse portal to provide tangible support for Australians of all ages who have had their intimate image or video shared without their consent. In September 2018, eSafety's powers in this area were expanded with the establishment of a civil penalties scheme.

As at mid-November 2019, eSafety had already received over 1800 reports of image-based abuse and 1100 this calendar year, which is an approximate 200% increase from the same time in 2018. Despite nearly all the websites that have been reported to date being hosted overseas, eSafety has been successful in having image-based abuse material removed in about 90% of cases where removal was requested.

eSafety's Cyber Report team investigates complaints from Australian residents and law enforcement agencies about harmful and illegal online content. eSafety prioritises

[13] Department of Communications and the Arts, Online Safety Legislative Reform: Discussion Paper, December 2019, pages 32 and 33

investigations into online child sexual abuse material, though may also investigate a range of other prohibited material, including material that: promotes matters of crime or violence; provides instruction in paedophilia; advocates terrorist acts; depicts gratuitous depictions of violence and sexual violence; or is sexually explicit. eSafety works with law enforcement and the global network INHOPE — the International Association of Internet Hotlines — to remove this content wherever it is hosted. The Cyber Report team is on track to complete 12,000 investigations in 2019, which is an approximate 50% increase from 2018.

eSafety's programs and initiatives

eSafety has a range of programs and initiatives in line with its comprehensive regulatory remit to promote online safety for all Australians.

Women

eSafety has a range of women's programs and initiatives.

eSafetyWomen

Launched in 2016 with funding under the Women's Safety Package, eSafetyWomen aims to empower all Australian women to manage technology risks and abuse. It helps women take control of their online experiences by:

- providing practical tools and information to equip women to protect themselves and their families against all forms of online abuse
- training frontline, specialist and support staff in the domestic violence sector, giving them the knowledge, skills and resources to effectively support women and their families, and
- actively raising awareness and understanding of technology-facilitated abuse to help women identify it and take steps towards preventing it.

A critical element of the eSafetyWomen program is the training program for domestic and family violence frontline workers through:

- delivery of intensive face-to-face workshops and customised presentations
- provision of eSafetyWomen—online training for frontline workers, and
- provision of web-based information and resources.

Earlier this year, eSafety released a range of fact sheets and 'top tip' postcards, covering the most important aspects of the eSafetyWomen program. Crucially, these were translated into 12 community languages.

Since its launch in June 2016, more than 10,000 domestic and family violence frontline workers have participated in face-to-face eSafetyWomen workshops.

Since July 2018, more than 2,200 have registered to undertake eSafetyWomen – online training for frontline workers.

Women Influencing Tech Spaces

eSafety's newest women's initiative, Women Influencing Tech Spaces (WITS), was launched in May 2018 to help women who are also disproportionately targeted for online abuse — those in the public eye. WITS is an initiative to both protect and promote women's voices. It recognises that women in leadership positions and with public personas, ranging from politics, business, media, sports and academia, experience shockingly high levels of abuse.

It also recognises, however, that social media can be, and should be, a powerful tool for women to engage, connect, communicate, learn and grow. One of the key resources developed for WITS was a suite of resilience tips and technique, which are designed to empower women will skills and strategies for managing cyber abuse.

eSafety's objective with WITS is to give women the psychological armour to counteract cyber abuse and interact online with impact, confidence and resilience.

Education and awareness

eSafety has a key leadership role in online safety education. It promotes, coordinates and leads online safety education for Australians nationally. This includes developing guidance and resources to assist schools in preventing and responding to online safety issues and undertaking research to identify best practice in online safety education.

Since 2015, and as at mid-November 2019, it has provided training to over 500,000 Australians, including:

- 13,641 teachers though professional learning sessions
- 10,952 pre-service teachers through learning sessions
- 306,872 students though - Virtual Classrooms
- 1180 chaplains to support the National Schools' Chaplaincy Program
- 28,229 community members
- 19,380 attendees at conferences, and
- another 68,802 Australians through other various sessions.

eSafety's education resources map to the Australian Curriculum and are promoted extensively through its education channels and stakeholders. Most recently, eSafety expanded its online safety resources for children, parents and teachers to include:

- The YeS Project (Young and eSafe) — a digital and social health program, which was launched in September 2018 and has been downloaded over 7,500 times.
- The Lost Summer — an educational role-playing video game, which was launched in July 2018 and has been downloaded over 28,000 times.
- Parent and Carer – a suite of new material that is focused on equipping parents and carers with strategies to help families deal with online issues, such as cyberbullying, online pornography, sending nudes and sexting, time online, gaming and unwanted contact.

Over 2020, eSafety will continue to develop the eSafety Early Years Program, the Online Safety Grants Program and the new Trusted eSafety Provider Program, which will ensure online safety education providers meet high quality standards and their programs are aligned to curriculum.

Seniors

eSafety's Digital Literacy for Older Australians program, Be Connected, is a joint initiative with the Department of Social Services. Building upon eSafety research that shows older Australians have the lowest levels of online participation of all Australians,[14] the Be Connected website offers a range of resources to help older Australians improve their digital literacy and stay safe online, including:

- activity based content for face-to-face and online learning
- tools for family members to encourage support and learning, and
- access to webinars and online safety workshops targeted at older Australians and those who work with them.

Since its launch in October 2017, the website has attracted 160,00 users and more than 300,000 learning activities have been completed.

[14] Office of the eSafety Commissioner, Digital Behaviours Understanding Report, September 2017

Raising awareness of support

Another concerning finding from eSafety's online hate speech research was that 64% of people who experienced hate speech took no action.

These findings closely mirror eSafety's research relating to children's cyberbullying and image-based abuse.

eSafety research released in May 2018, which was based on a National Youth Survey, showed that only around 24% of young people who had a negative online experience sought help in a formal way through their school, a social media company or the police.[15]

This research showed that barriers to young people reporting cyberbullying include feelings of shame and embarrassment, fear of retaliation and fear of not being believed. Young people also had fears about making a complaint, including that they would lose control over the complaints process and that the issue would escalate and become public, resulting in the loss of their anonymity.

eSafety's research with female victims of image-based abuse showed that most women did not take formal action in seeking help and support from professional services, such as the police, support services or legal advice.[16]

This is incredibly concerning, given the harms victims can suffer. It is also concerning because reporting cyber abuse is one of eSafety's most important safety recommendations. Through the feedback processes of its complaint services, eSafety knows that reporting cyber abuse can both provide a sense of relief and be empowering for the victim, as it helps restore to them a sense of control. Having harmful content taken down also helps de-escalate the trauma and minimise the possibility of re-traumatisation, which may be heightened if the material remains online.

Reporting also puts the platforms on notice about the nature, volume and trends of abuse on their platforms. Over time, this should help improve the challenges posed by online abuse at an institutional and platform level.

This reinforces the need to raise awareness about reporting services, including the unique services of eSafety and anti-discrimination bodies that often handle anti-vilification complaints, including the Victorian Equal Opportunity and Human Rights Commission.

[15] Office of the eSafety Commissioner, State of Play — Youths, Kids and Digital Dangers, May 2018
[16] Office of the eSafety Commissioner, Image-based abuse: National Survey - Summary Report, October 2017

It also reinforces how eSafety's services and anti-vilification protections work in a complementary and mutually reinforcing manner.

eSafety recommends the Committee promote the wide-ranging services of eSafety in formulating a multi-faceted and holistic approach to addressing online harm, which both includes and is broader than online vilification.

Cooperative arrangements with social media companies

eSafety's regulatory success is founded on its cooperative approach with social media companies.

While the eSafety Commissioner has a range of regulatory powers and penalties to compel compliance with its formal legislative schemes, eSafety prefers to adopt a cooperative approach. Given the strong working relationships eSafety has built with the major social media services, to date it has achieved a 100% compliance rate with its request to platforms to remove child cyberbullying content, often within a matter of hours, without needing to use its formal powers.

However, eSafety also notes that an important reason this cooperative arrangement has been successful is precisely because eSafety has such regulatory powers at its disposal. This links to the important role eSafety adopts as an advocate on behalf of the user, as there is often an inherent power imbalance between the person being abused and the corporate social media service. It is therefore a balance of regulatory settings that is key.

In addition to its regulatory schemes, another example of eSafety's successful cooperative approach is Safety by Design (SbD).

Safety by Design

Educating and empowering people will always form the basis for addressing the social and behavioural issues that manifest online. However, technological approaches, interventions and frameworks also play a critical role in reducing online harm.

Indeed, the same technology platforms that can be used to spread positivity and ideas can be used for more nefarious purposes, including cyber abuse and online vilification.

In June 2018, eSafety laid out its intention to develop a SbD Framework and set of voluntary SbD Principles. At its core, SbD is about embedding the rights of users and user safety into the design, development and deployment of online and digital products and services. It seeks to ensure user safety is proactively considered as a standard risk mitigation and development process, rather than retrofitted after online harms have occurred.

SbD places user safety considerations at the centre of product development. It recognises and responds to the intersectionality of risk and harm in the online world and acknowledges the potential of advancements in technology, machine-learning and artificial intelligence to radically transform user safety and online experiences.

Following an eight-month consultation process with industry, parents and carers and children, eSafety developed three overarching SbD Principles: service provider responsibilities; user empowerment and autonomy; and transparency and accountability.

eSafety is currently embarking on phase two of the SbD initiative, which is aimed at creating a Framework of guidance and resources to help facilitate the adoption of the Principles by various industry sectors. eSafety will be paying particular attention to developing guidance for small and medium size enterprises, and the start-up community, as well as generating opportunities for collaboration and the sharing of safety-enhancing tools, best practices and technologies.

In securing a more ethical, value-centred and human-centred approach to the development of technologies, SbD can act as a catalyst for further innovation and lead to lasting cultural change within industry.

Conclusion

As the world’s first and only government agency dedicated solely to online safety, eSafety knows that online issues are only becoming more complex, pervasive and challenging.

This underscores the important role eSafety plays in leading, coordinating and advising on online safety issues to ensure all Australians have safe, positive and empowering experiences online. It also underscores the importance of online vilification laws that are commensurate in scope and nature to the intersectional nature of cyber abuse, which is regrettably common in contemporary society and extends beyond race and religion.

eSafety hopes its submission has highlighted the importance of broad vilification laws, as well as the important role it plays in protecting Australians online, as part of a multi-faceted and holistic approach to addressing online harms.