Australian Government
Department of Communications and the Arts

Online Safety Legislative Reform

Discussion paper

DECEMBER 2019

Disclaimer

The material in this paper is of a general nature and should not be regarded as legal advice or relied on for assistance in any particular circumstance or emergency situation. In any important matter, you should seek appropriate independent professional advice in relation to your own circumstances. The Commonwealth accepts no responsibility or liability for any damage, loss or expense incurred as a result of the reliance on information contained in this paper.

This paper been prepared for consultation purposes only and does not indicate the Commonwealth's commitment to a particular course of action. Additionally, any third party views or recommendations included in this paper do not reflect the views of the Commonwealth, or indicate its commitment to a particular course of action.

Copyright

Third party copyright

Attribution

Using the Commonwealth Coat of Arms

Contents

Executive summary

This discussion paper outlines the key elements of a proposed new Online Safety Act and seeks comments and views on these proposals. Keeping Australians safe online is a key policy priority for the Government and this program of legislative reform is pivotal in achieving this goal.

The proposed new Act would bring together the separate components of the existing online safety regulatory framework in a single place. It would build on the strengths of the existing schemes regulating cyberbullying and image-based abuse in the *Enhancing Online Safety Act 2015* and extend these schemes. It would migrate elements of the current online content scheme into the new Act, spark the creation of new industry codes to address harmful online content, and allow the eSafety Commissioner to address the most harmful content wherever it is hosted.

As far as possible, the revised online safety schemes would be applied consistently to the different types of online service providers, and not just large social media companies and Australian internet service providers (ISPs).

The Act would enable the development of a set of new basic online safety expectations to make clear that the Government expects online service providers to deliver improved online safety outcomes. The eSafety Commissioner would also be provided with the power to mandate companies to provide transparency reports on how these expectations are being met.

The new Act would create a new cyber abuse scheme for Australian adults to tackle the most serious forms of online abuse. The eSafety Commissioner has had great success in working with social media companies to remove material in very short time frames – even as short as 30 minutes. It is proposed that the time to respond to take-down notices under all four online safety schemes be shortened from 48 to 24 hours.

The new Act would include most elements of the online content scheme currently found in the Broadcasting Services Act 1992. However, the scheme would be expanded to empower the eSafety Commissioner to determine whether particular online content was harmful enough to require take-down action, rather than the current time-consuming requirement for prohibited online material to be assessed and classified by the Classification Board.

It would also establish clear and unambiguous power for the eSafety Commissioner to protect Australians during an online crisis event (similar to the mass shootings that occurred in Christchurch in March 2019) by directing ISPs to block access to sites hosting terrorist and violent material.

The new Act would introduce a new ancillary service provider notice scheme for parts of the online services sector that are not directly responsible for harmful content, but that provide access to it through their services. The eSafety Commissioner could ask these services to support demotion, de-ranking or removal.

Over the past four years, the responsibilities of the eSafety Commissioner have expanded in scope and scale. The new Act would clarify the functions and governance arrangements for the Office of the eSafety Commissioner, which currently operates with the support of another agency, the Australian Communications and Media Authority (ACMA).

The following figures provide an outline of the current online safety legislative framework and the proposed components of a new Online Safety Act.

Figure 1: Outline of the existing online safety legislative framework

Enhancing Online Safety Act 2015			Broadcasting Services Act 1992
Promoting online safety for all Australians	Complaints and take-down system for cyberbullying of Australian children	Complaints and take-down system for image-based abuse	Complaints and take-down system for prohibited and potential prohibited online content

Figure 2: Proposed outline of the new Online Safety Act

Online Safety Act			
Object	Promote online safety	Prevent online harms	Protect Australians online
Functions	Education	Basic online safety expectations	Cyberbullying of children
	Coordination		Cyber abuse of adults
	Evaluation	Opt-in tools and services	Image-based abuse
	Research		Seriously harmful material
	Grants		Blocking in online crisis events

Introduction

The Government intends to reform the existing online safety legislative framework by developing a new Online Safety Act. This discussion paper is designed to inform a process of consultation on the elements of a new Act and seeks views on the possible elements and the impacts of various options.

Over the past two decades, Australia has been at the forefront of online safety policy and regulation. In 1999 it extended the broadcasting regime to deal with harmful online content, including child sexual abuse material.[1] In 2015, the Government established the world's first Children's eSafety Commissioner to address the particular harms faced by children online. This became the eSafety Commissioner in 2017 when the remit was extended to include all Australians.

Over time, it has become clear that some elements of the legislation underpinning the online safety regime are out of date and not flexible enough to deal with new and emerging issues. These arrangements were tested in the aftermath of the Christchurch mass shootings in March 2019 and demonstrated that they were not up to the task of timely regulatory intervention to address the viral proliferation of the footage of the shootings and the alleged perpetrator's manifesto.

An independent review of the legislation in 2018 recommended that there be a single up to date Online Safety Act. This would allow key elements of the legislative framework to be modernised and improved. In particular, the reliance on the National Classification Scheme in the online content scheme is administratively cumbersome. Following that review, the Government made a commitment in the context of the 2019 election to develop and implement a new Act.

Objectives for reform

In consulting on the development of a new Act, the Government is seeking to fulfil a range of objectives. These are to:

1. maintain the elements of the existing framework that are working well, such as the cyberbullying and image-based abuse schemes;
2. address gaps in current regulatory arrangements, particularly where the current schemes are out of date or don't address harms occurring on more recently developed services and platforms;
3. establish a more flexible framework that can accommodate new online harms as they emerge;
4. hold the perpetrators of harmful online conduct to account for their actions online;
5. improve the transparency and accountability of online service providers for the safety of their users and the mitigation of online harms;

[1] Commonwealth law uses the term 'child abuse material' to capture material that depicts or represents the sexual or physical abuse of a person who is or appears to be under 18 years of age.

6. enable the eSafety Commissioner to continue to protect Australians online, promote online safety and prevent online harms; and
7. provide all Australians with the information, tools and resources necessary to engage safely online and build resilience to potential online harms.

What is online safety?

'Online safety' is used throughout this paper to refer specifically to the mitigation of harms that can affect people through exposure to illegal or inappropriate online content or harmful conduct. It does not encompass the full range of risks that Australians face online. It does not address cyber security threats to the privacy, availability and integrity of Australians' data and networks. It also does not refer to harms caused by online gambling and the promotion of gambling content online. These potential online safety risks are dealt with by other legislative regimes.

Concurrent reviews

The Government has recently agreed to conduct an independent review of the National Classification Scheme which provides a framework by which films, computer games and certain publications made available in Australia receive a rating and consumer advice that provides a safeguard to the Australian public that content is consumed by the appropriate audience. That review will run in parallel to the development of a new Online Safety Act.

In June 2019, the Australian Competition and Consumer Commission (ACCC) completed an inquiry into digital platforms with recommendations relating to improving consumer outcomes by changes to regulation about online advertising, privacy and the use of data. The Government response to these issues is progressing separately to this process.

The Government recently completed an initial consultation to inform the development of the 2020 *Cyber Security Strategy*. This included calling for submissions in response to a public discussion paper. This separate body of work to review Australia's cyber security arrangements is complementary to the reform of online safety legislation.

The Commonwealth Attorney-General announced in November 2019 that the Council of Attorneys-General would be considering progressing major reform in defamation laws, informed by public consultation.

How to make a submission

The Department is seeking views from stakeholders and interested parties in response to the options and the questions put forward in this discussion paper.

The Department invites submissions by **5.00 pm AEDT on Wednesday, 19 February 2020**. Submissions may be lodged in the following ways:

Website	www.communications.gov.au/have-your-say
Email	onlinesafety@communications.gov.au
Post	Director, Online Safety Research and Reform Section Department of Communications and the Arts GPO Box 2154 Canberra ACT 2601

Submissions should include your name, organisation (if relevant) and contact details. The Department will not consider submissions without verifiable contact details.

Submissions will be treated as non-confidential information, and will be made publicly available on the Department's website unless you specifically request that your submission, or a part of a submission, be kept confidential, and provide acceptable reasons. An email disclaimer asserting confidentiality of the entire submission is not sufficient, nor is a header or footer disclaimer.

The Department reserves the right not to publish a submission, or any part of a submission, at its absolute discretion. The Department will not enter into any correspondence with respondents in relation to any decisions not to publish a submission in whole or in part.

The Department is subject to the *Freedom of Information Act 1982* and may be required to disclose submissions in response to requests made under that Act.

The *Privacy Act 1988* establishes certain principles regarding the collection, use and disclosure of information about individuals. Any personal information respondents provide to the Department through submissions will be used for purposes related to considering issues raised in this paper, in accordance with the Privacy Act. If the Department makes a submission, or part of a submission, publicly available, the name of the respondent will be included. Respondents should clearly indicate in their submissions if they do not wish their name to be included in any publication relating to the consultation that the Department may publish.

Questions about the submission process can be directed to onlinesafety@communications.gov.au

Next steps

This consultation process will inform the Government's development of a new Online Safety Act. In particular, information about the impact of particular options on businesses and the community would be beneficial in developing the legislation.

Background

Existing online safety framework

The existing online safety framework focusses on the mitigation of the range of harms that can affect people based on exposure to illegal or inappropriate online content or harmful conduct. These arrangements are set out in the *Enhancing Online Safety Act 2015* (EOSA) and Schedules 5 and 7 of the *Broadcasting Services Act 1992* (the online content scheme).

The EOSA establishes the powers and functions of the eSafety Commissioner, an independent statutory office holder that currently operates with the support of the ACMA.

The functions of the eSafety Commissioner include education, coordination, grants administration and research in relation to online safety, as well as the oversight of three regulatory schemes:

- the cyberbullying scheme that addresses cyberbullying of an Australian child;
- the image-based abuse scheme that addresses the non-consensual sharing of intimate images; and
- the online content scheme for regulating access to illegal and harmful online content ('prohibited content' and 'potential prohibited content').

Under these three schemes, the eSafety Commissioner acts in response to complaints about material along with some capacity to initiate investigations relating to online content. The eSafety Commissioner works with social media platforms, content hosts and internet service providers to have content removed on a voluntary basis. This approach has been successful in the vast majority of cases in relation to Australian-hosted internet content, cyberbullying material and intimate images.

The eSafety Commissioner is also able to take enforcement action under the three schemes (with the exception of overseas hosted content). If content is hosted overseas, there is no power in the online content scheme to issue take-down notices. However, the eSafety Commissioner can and frequently does report particularly serious content, such as child sexual abuse material, to international law enforcement for investigation and removal. The eSafety Commissioner is also able to trigger the blocking of access to certain overseas hosted content through notifying Australian internet service providers of that content.

A detailed outline of the three online safety schemes, and the role and functions of the eSafety Commissioner, is provided in the discussion of a proposed new Online Safety Act in Section 5.

In addition to the powers in the three online safety schemes, in April 2019 the eSafety Commissioner was given new powers by the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* to issue notices to content service providers and hosting service providers about the presence of abhorrent violent material (AVM). If a service provider does not act to remove AVM, it could be subject to prosecution through criminal law.

Review of online safety legislation

In 2018, the Government commissioned an independent review of the online safety legislative framework by Ms Lynelle Briggs AO.[2] This review found that the existing framework has worked reasonably well, despite some gaps in coverage, but reform was needed to make the framework fit for purpose and able to deal with new and emerging technologies. The review recommended:

- replacing the existing legislation with a single Act;
- increasing the expectations on online service providers to be proactive in preventing online harms;
- extending the cyberbullying scheme to include material directed towards adults; and
- changing the governance arrangements of the Office of the eSafety Commissioner to address limitations and deficiencies in the current arrangements.

The proposals canvassed in this discussion paper respond to, and have been informed by, the findings and recommendations of the 2018 Review.

Election commitments

The Government is committed to improving online safety for all Australians. In this context, the Government made a number of commitments to improve online safety in the lead up to the 2019 federal election. These include commitments to undertake legislative reform, to collaborate with online service providers, and to pursue domestic and international initiatives. Specifically, the Government has committed to the following actions and initiatives.

A new Online Safety Act

The Government has committed to the development of a new Online Safety Act to help protect Australians online. The development of a new Act presents an opportunity to consolidate existing regulatory arrangements and to update them in light of changes in the digital media environment. This discussion paper is an important step in the development of the new Act.

Requiring stronger privacy settings for devices and services marketed to children

Games and apps are being used by Australian children and, in some cases, they can be conduits for offensive and harmful content and conduct. The Government has committed to working with online service providers to make online apps, games and services marketed to children default to the most restrictive privacy and safety

[2] Lynelle Briggs 2018 *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme)*

settings at initial use or set-up. The community expects that online safety principles, protections and processes should be embedded into the design of games and apps that prevent them from being used by online predators. The Government has articulated these expectations through its Online Safety Charter (discussed below) and, if necessary, will legislate to codify requirements for online service providers.

Availability of a filtered internet service to keep children safe online

Parents and carers are concerned with the ease with which children can access disturbing content or pornography online – either deliberately or by accident. Preventing children's access to harmful content requires a mixture of approaches including regulation, education and building resilience. Technology can also play a role and as such, the Government has committed to working with ISPs to make available to parents and carers the option of a filtered internet service that, at a minimum, blocks access to websites identified by the eSafety Commissioner.

Making online safety information available at the point of purchase

Parents and carers need information to make informed choices about the devices, apps and games their children use. While the eSafety Commissioner provides an eSafety Guide that includes online safety information about a range of games and apps, and advice on parental controls and tools to minimise the risks associated with the use of devices by children, more information should be available at the point of purchase or download. As such, the Government has committed to working with retailers and service providers to provide information regarding online safety and parental controls at all points in the supply chain including point-of-purchase, registration, account creation and first use.

Mandating transparency from major social media platforms

Policy responses to online harms must be tailored to the services in question and calibrated to the scale and impact of harms that are occurring. Social media platforms maintain user complaint and moderation processes to address content that breaches their terms of service and use. As a result, social media platforms have access to data about user reports of harmful content, the volume of content found to breach policies, and the actions taken. However, the reports issued by social media platforms vary in terms of quality and the level of detail provided. As a result, the Government has committed to mandating transparency reports, aligned with international efforts, from social media companies. Reports would provide data on the number and type of responses to reports and complaints about illegal, abusive and predatory content by users.

Working with the international community

The online world is not contained by geographic boundaries and as such, addressing harmful online content requires a global response. The Government has committed to working with G20 Leaders to remain engaged with technology firms' progress to meet obligations regarding prevention and protection, transparency and deterrence. This commitment has been achieved by securing a G20 Leaders' Statement at Osaka, Japan, announced on 29 June 2019, on preventing exploitation of the internet for terrorism and violent extremism. Australia will continue to advocate strongly in a range of international forums, including the OECD and through other bilateral relationships, for a safer online environment.

A higher bar for online service providers

The eSafety Commissioner's Safety by Design principles

During 2018 and 2019, the eSafety Commissioner developed a set of voluntary Safety by Design (SbD) principles to place the safety and rights of users at the centre of the design, development and deployment of online products and services.

The eSafety Commissioner consulted widely with online service providers and users, including young people, on the principles and has promoted the SbD concept around the world. Google, Facebook and Snap are among the major companies that have agreed to take on the SbD principles and use them to develop, design and deploy new technology.

The eSafety Commissioner is continuing to work closely with online service providers to progress implementation of the SbD principles, a process that will include the development of an implementation guide for online service providers to guide systemic shifts in the design of services.

Online Safety Charter

The Australian Government has released its Online Safety Charter.[3] The Charter articulates the Government's expectations, on behalf of the Australian community, of the steps online service providers should take to protect their users, especially children and vulnerable members of the community, from harmful online experiences.

The Charter is based on the premise that behaviour that is unacceptable offline should not be tolerated or enabled online. It acknowledges that online providers have a responsibility to take meaningful action to address and prevent harms from being incurred by those using their products or services.

[3] Available at www.communications.gov.au

The Charter endorses the eSafety Commissioner's SbD principles as best practice, expanding on those expectations in some areas of particular importance. The Charter includes the following expectations:

- **Service provider responsibilities** – preventative steps that service providers should take to reduce the potential for their services to facilitate, inflame or encourage illegal and inappropriate behaviours.
- **User empowerment and autonomy** – the measures, tools, mechanisms, protocols, policies, features and practices that service providers and technology firms should have in place to empower users to enjoy safe online interactions.
- **Transparency and accountability** – the provision of information to employees, users, researchers, civil society and governments on online safety metrics to inform the assessment and development of improved online safety outcomes across the sector.

The Charter is a statement of expectations rather than, at this point, formal legal requirements with sanctions attached for non-compliance. However, the SbD principles, the Charter, and the outcomes of this consultation process will inform the drafting of the proposed basic online safety expectations for inclusion in the Online Safety Act. This proposal is discussed further in section 5 of this paper.

Online safety challenges and the need for reform

Current issues

Online interaction is pervasive in Australian life and as such, online safety is an important issue for all Australians. Online interactions permeate all aspects of modern life: to work; to socialise; to consume; and to engage with government, education, health and financial systems.

- In the six months to May 2018, 89 per cent of Australian adults had accessed the internet, with universal access by those aged 18 to 44 (100 per cent).
- As at May 2018, 74 per cent of online Australian adults had been active on social media sites in the last six months.
- Australian adults also participate in a diverse range of online activities: sending and receiving email (95 per cent); researching or gathering information (94 per cent); and general internet browsing (93 per cent) being the most popular activities as at May 2018.
- Four in five (82 per cent) Australian internet users viewed video content online in the six months to May 2018, while three in five (61 per cent) accessed audio content such as internet radio or podcasts.[4]

Online engagement also impacts Australians at all life stages — from pre-school aged children to the elderly.

- In 2018, 90 per cent of children aged 5 to 14 years were looking at screens each week, most for ten hours or more.[5]
- 81 per cent of parents with pre-schoolers aged 2 to 5 years said their children were using the internet in 2018, and 99 per cent of parents with children aged 2 to 17 years reported having an internet connection in the home.[6]
- Six per cent of preschoolers were using social media and 20 per cent were playing multi-player online games in 2018.[7]
- Australians aged 50 to 69 are significantly more engaged with the internet than their older counterparts. Those who are 70+ years old cited a lack of trust, confidence, skills and personal relevance for their digital disengagement. The study, conducted in 2017, also found that older Australians have fears about going online, particularly related to security, with close to half surveyed reporting experiences related to virus, scam, credit card and personal information theft.[8]

[4] ACMA, *Communications report 2017–18* (2019)

[5] ABS media release 26 March 2019 'Kids clock up 10 or more hours of screen time per week' https://www.abs.gov.au/ausstats/abs%40.nsf/Latestproducts/4921.0Media%20Release202017-18?opendocument&tabname=Summary&prodno=4921.0&issue=2017-18&num=&view

[6] eSafety Commissioner, *Digital Parenting – supervising children online* (2018), www.esafety.gov.au/about-us/research/digital-parenting/supervising-preschoolers-online www.esafety.gov.au/about-us/research/digital-parenting/digital-families

[7] eSafety Commissioner, *Digital Parenting – supervising children online* (2018).

[8] eSafety Commissioner, *Understanding the digital behaviours of older Australians* (2018).

The internet is an integral part of the digital lives of young people in Australia, with most going online regularly to learn, keep in touch with friends and have fun. However, key themes found by the eSafety Commissioner in relation to the online challenges facing young people age 8–17 include:

› They are exposed to a wide range of issues online, including unwanted contact and bullying, and deal with these issues in a range of ways.

› While negative experiences can be hurtful, young people also report positive outcomes from these experiences in terms of increased awareness of online risks and ways of dealing with issues when they arise.

› Young people are not alone in having to deal with the unpleasant aspects of online participation with adults also experiencing similar challenges. This is a reflection of the importance of ongoing learning to build digital resilience and respect online.[9]

The eSafety Commissioner's extensive consultation on the SbD principles included engagement with young Australians, revealing their expectations of online service providers. These included: the need for a strong set of easy-to-understand and highly visible ground rules that have user safety at their core; and an expectation that online service providers are aware of the issues faced by users, and are responsible for promoting their safety by prioritising them above all else.

Online harms are complex and continue to evolve

Online safety issues can be complex, and are often an online extension of pre-existing social issues, such as family and intimate partner violence and face-to-face bullying or stalking. Online harms include cyberbullying, abusive commentary or 'trolling', the non-consensual sharing of intimate images (image-based abuse), grooming for the purpose of child sexual abuse, cyberflashing, doxing and cyberstalking. Online safety measures extend to mitigating user exposure to illegal or harmful content, such as extremely violent content, terrorist propaganda or child sexual abuse and exploitation material.

New forms of online harms have emerged globally as services, businesses, education and social interactions become increasingly digitised and connected. These recent technological developments have presented new and significant benefits, but they have also presented new regulatory challenges, especially on the front of addressing technology-facilitated crime and abuse.

This includes new sources of online harm that were not envisaged when the current regulatory framework was first developed. For example, online multi-player gaming services were not widely available in 1999, when the online content scheme was first developed, and were still relatively immature in terms of market penetration in 2007 when the scheme was expanded to cover a wider range of online services.

[9] eSafety Commissioner, *State of Play – Youth, kids and digital dangers* (2018).

Yet, in 2018, research into youth and gaming found that 17 per cent of multiple player gamers experienced in-game bullying.[10]

There are also increasing numbers of people reporting online harms. Since 2015, the eSafety Commissioner has received 1,547 cyberbullying complaints, with 51,923 visits to their complaints form. 531 of these reports were in 2018-19, a 30 percent increase on the previous year.

Online safety is increasingly seen as a key issue internationally, with recognition that online harms have no borders. The response requires a multi-pronged approach addressing prevention, protection and response, and effective coordination of national and global efforts.

Online harms give rise to significant costs

Unchecked online harms present a variety of risks to Australians.

Negative online experiences can exacerbate social exclusion and psychological harm. There is increasing evidence that both face-to-face bullying and cyberbullying have lasting effects on young people, including poor self-esteem and mental health, depression, anxiety and suicidal ideation.[11] A 2018 review of studies of cyberbullying, self-harm and suicidal behaviour amongst children and young people published between 1996 and 2017 found that having been a victim or perpetrator of cyberbullying is associated with significantly higher rates of self-harm or attempted suicide than for non-victims and non-perpetrators.[12]

The consequences of cyberbullying are often felt well beyond the perpetrator and victim involved, impacting families, friends and local communities. Schools are often adversely impacted, as are service providers such as out of home care organisations. Online harms can negatively affect social cohesion in Australia. For example, terrorist and violent extremist content can aggravate tensions, spread fear and be used to radicalise at-risk individuals.[13]

Online safety issues can also have economic effects when victims reduce their participation in the workforce or need medical or psychological help. A 2018 survey commissioned from the Australia Institute found that 39 per cent of adult internet users reported receiving online harassment. Four per cent of respondents reported that their ability to work was impaired, or reported seeing a doctor, psychologist or health professional as a result of harassment.

[10] https://www.esafety.gov.au/-/media/cesc/documents/corporate-office/youth_and_gaming_doc.docx

[11] Australian Government Department of Education and Training, Submission 2 to the Senate Committee Inquiry on the Adequacy of existing offences in the Commonwealth Criminal Code and of state and territory criminal laws to capture cyberbullying p. 4.

[12] John A, Glendenning AC, Marchant A, Montgomery P, Stewart A, Wood S, Lloyd K, Hawton K 'Self-Harm, Suicidal Behaviours, and Cyberbullying in Children and Young People: Systematic Review' *J Med Internet Res* 2018;20(4):e129 URL: https://www.jmir.org/2018/4/e129

[13] Cases of online radicalisation in Australia include Jake Bilardi: The radicalisation of an Australian teen see https://www.bbc.com/news/world-australia-31845428

Under the most conservative estimate, online harassment and cyber-hate were estimated to have resulted in $62 million in medical costs and $267 million in lost income for Australians.[14] The Australia Institute projected the economic costs across the population to be between $330 million and $3.7 billion to date.[15] More research would be needed to develop a longitudinal estimate of the economic impacts each year.

Research undertaken by PwC in 2018 investigated the economic cost of bullying in Australian schools and estimated that these costs totalled $2.3 billion, incurred while the children are in school and for 20 years after school completion, for each individual school year group.[16]

While major online service providers continue to invest in the safety of their services, the approaches taken by different companies are not consistent, and community members continue to call for stronger preventative measures – as demonstrated through the eSafety Commissioner's engagement with young people on SbD,[17] submissions to the 2018 Review,[18] and the public consultation process informing the Government's development of the Online Safety Charter.

Global regulatory responses

A number of overseas jurisdictions have also taken proactive measures to improve the accountability of digital platforms for content and behaviour on their services.

- In the United Kingdom, the Government is currently developing legislation to implement the measures in its **White Paper on Online Harms**. This recommends a new statutory duty of care that will legally oblige social media platforms to take reasonable and proportionate steps to stop and prevent harmful material appearing online. There would be an independent regulator with strong enforcement powers to deal with non-compliance.[19] However, in October 2019, the UK Government announced that it not be proceeding with its proposed mandatory age verification scheme for access to online pornography by adults but would instead give a regulator discretion as to the most effective means for companies to meet their duty of care.[20]

[14] The Australia Institute 'Trolls and Polls – the economic costs of online harassment and cyberhate', January 2019

[15] ibid.

[16] PwC 'The economic cost of bullying in Australian schools' 2018 https://www.amf.org.au/media/2505/amf-report-280218-final.pdf

[17] https://www.esafety.gov.au/esafety-information/-/media/cesc/sbd/safety_by_design_overview.pdf

[18] Published at https://www.communications.gov.au/publications/report-statutory-review-enhancing-online-safety-act-2015-and-review-schedules-5-and-7-broadcasting

[19] UK Online Harms White Paper, released 8 April 2019: www.gov.uk/government/news/uk-to-introduce-world-first-online-safety-laws.

[20] Written Ministerial Statement of the Secretary of State for the Department for Digital, Culture, Media and Sport, Nicky Morgan, 16 October 2019: https://www.parliament.uk/business/publications/written-questions-answers-statements/written-statements/

- In Germany, the Netzwerkdurchsetzungsgesetz (**Network Enforcement Act**) requires internet platforms with more than 2 million users to have in place reporting systems for hateful posts and to delete reported content if it is illegal under the German Criminal Code within 24 hours. Implemented in 2018, this legislation has reportedly led to Facebook increasing the German based staff resources dedicated to moderating German content.[21]
- In Europe, the European Commission has developed a **Code of Conduct on Countering Illegal Hate Speech Online**. Since May 2016, Facebook, Twitter, YouTube and Microsoft have committed to combatting the spread of racist and xenophobic content and terrorist propaganda in Europe through this code. Other platforms more recently announced they will participate under the Code, including Instagram and Google+ (January 2018), Snapchat (May 2018) and Dailymotion (June 2018).The European Union has indicated that it will consider additional measures, including legislative measures, if efforts to implement the Code are not pursued or slow down.[22]
- In Canada, the government launched a **Digital Charter** of principles in May 2019 that that Canadian people could expect, including that digital platforms will not foster or disseminate hate, violent extremism or criminal content.[23]
- In France, the French Parliament passed legislation in their lower house in July 2019 requiring online platforms to remove 'overtly hateful' content within 24 hours or face fines of up to €1.25m.[24] In May 2019, the French Government released an interim report (*Creating a French framework to make social media platforms more accountable: Acting in France with a European vision*) following a '**Facebook experiment**', in which French Government officials were embedded within Facebook, to inform the regulation of social networks.[25] This report proposed to address the issue of the borderless internet by requiring that platforms be accountable to the destination country (where a harm occurs), rather than the country in where the digital platform may be based.

[21] UK Digital, Culture, Media and Sport Committee's Disinformation and 'fake news' final report February 2019, pp.12-13 https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf

[22] European Commission 'Countering illegal hate speech online #NoPlace4Hate' https://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=54300

[23] Canada's Digital Charter: Trust in a digital world: https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00108.html

[24] Proposition de Loi visant à lutter contre la haine sur internet, http://www.assemblee-nationale.fr/15/propositions/pion1785.asp

[25] https://www.france24.com/en/20190510-france-facebook-law-mark-zuckerberg-president-macron-internet-regulation-internet

Case Study: Global response to the Christchurch attacks

The terrorist attacks that took place in Christchurch on 15 March 2019 shocked the world, as the mass murder of 51 men, women and children was live-streamed and the resulting footage distributed globally. The attacks provided the impetus for greater government and online service sector efforts to improve online safety and prevent the misuse to the internet for terrorist purposes.

- In Australia, the Government established a joint government industry **Taskforce to Combat Terrorist and Extreme Violent Material Online** to develop short and medium term recommendations to prevent the recurrence of this use of the internet for terrorist purposes. The Taskforce produced a consensus report recommending a range of actions, now being implemented, for government and industry to improve their ability to prevent and respond to future online crisis events. One of these recommendations was to establish a clear content blocking framework for terrorist and extreme violent material online in future crisis events.[26] This recommendation is incorporated in this discussion paper.
- The Government also developed amendments to the *Criminal Code Act 1995* to give the eSafety Commissioner power to issue notices to advise content and hosting service providers of the presence of **abhorrent violent material** (AVM) on their services. These amendments introduced a new offence for content service providers and hosting services if they do not ensure the expeditious removal of, or cease hosting, the AVM. The amendments also included a new offence for service providers which fail to report abhorrent violent material that records or streams abhorrent violent conduct that is occurring in Australia to the Australian Federal Police within a reasonable timeframe.[27]
- On 15 May, Australia joined the **Christchurch Call to Action** founded by New Zealand and France to bring together governments and online service providers to eliminate terrorist and violent extremist content online through voluntary measures.[28] Online service sector participants who also supported the Call (Twitter, Amazon, Google, Microsoft and Facebook) issued a set of nine actions to address the abuse of technology to spread terrorist and extremist content.[29]

[26] Report of the Australian Taskforce to combat terrorist and extreme violent material online, 30 June 2019 https://www.pmc.gov.au/resource-centre/national-security/report-australian-taskforce-combat-terrorist-and-extreme-violent-material-online

[27] *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*

[28] https://www.christchurchcall.com/call.html

[29] https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2019/05/Christchurch-Call-and-Nine-Steps.pdf

- In June at a meeting of the G20 in Osaka, Japan, leaders adopted a statement proposed by Australia calling on governments and industry to work together to step up efforts to tackle terrorist and violent extremist exploitation of the internet. The **G20 Leaders' Statement** offered strong condemnation of terrorism and violent extremism in all its forms, and provides firm political pressure — the first with the support of the full G20 membership — to step up efforts to tackle terrorist and violent extremist exploitation of the internet. This Statement committed leaders to remain engaged with industry and to share experiences through international fora and initiatives.[30]
- In August 2019, the Australian Government announced that it, along with New Zealand and the OECD, would fund a project to develop a **voluntary transparency reporting protocol** for online platforms. This project will develop a common global standard for online platforms to report on the steps they take to prevent, detect and remove terrorist and extreme violent content. This protocol is expected to be developed during 2020.[31]
- The **Global Internet Forum to Counter Terrorism** (GIFCT) was established in 2017 as a small industry body for social media companies to share information about illegal content to streamline the processes for removing it. In September 2019, the founding companies announced that as part of implementing the Christchurch Call, the GIFCT would be strengthened and become an independent organisation with teams assigned to work on technology, counter-terrorism and operational issues.[32]

The Christchurch attacks, and the actions that have been taken since this tragic event, demonstrate the global nature of contemporary online safety threats. In response to new and rapidly changing threats, governments, industry and others are using new approaches to try to tackle them.

[30] https://g20.org/en/documents/final_g20_statement_on_preventing_terrorist_and_vect.html

[31] https://www.pm.gov.au/media/more-action-prevent-online-terror

[32] https://gifct.org/press/next-steps-gifct/

Components of proposed new Act

This section outlines and seeks comment on the elements of the proposed new Online Safety Act.[33] The key components of the proposed new Act include:

- the **objects** and **statement of regulatory policy** for the new Act;
- extending the **cyberbullying** scheme to more types of services;
- introducing a **new cyber abuse scheme for adults**;
- shortening the take-down time from 48 to 24 hours for both the cyberbullying and **image-based abuse schemes**;
- the introduction of a statement of **basic online safety expectations**;
- the current **online content scheme** to address illegal and harmful content with some expanded powers for the eSafety Commissioner;
- opt-in tools to restrict access to content that is **inappropriate for children**;
- empowering the eSafety Commissioner to implement **targeted blocks of terrorist or extreme violent material** during an online crisis event; and
- a scheme to reduce the availability of harmful material on **ancillary service providers**, such as search engines and app stores.

Objects of the new Act

Current approach

Unlike other pieces of legislation in the communications portfolio, the current EOSA does not have an objects section. For example, the *Telecommunications Act 1997* includes a comprehensive set of objectives at section 3 and a statement of regulatory policy at section 4. The purpose of an objects section is to set out the underlying purposes for a piece of legislation which can be used to aid interpretation of detailed provisions, including by the courts.[34]

Proposal

It is proposed that a new Online Safety Act, include a set of high level objects of:

- preventing online harms;
- promoting online safety; and
- protecting Australians online.

[33] The precise construction of provisions will be a matter for the Office of Parliamentary Counsel as part of the normal processes for drafting legislation.

[34] See section 15AA of the *Acts Interpretation Act 1901 (Cth)*, which provides when interpreting a provision 'the interpretation that would best achieve the purpose or object of the Act (whether or not that purpose or object is expressly stated in the Act) is to be preferred to each other interpretation.'

To support these broad objects, it is proposed that the Act include a statement of regulatory policy. This would articulate the intended outcomes of the Act and sit underneath, and elaborate on, the objects of the Act. For the purposes of this consultation process, the proposed statement of regulatory policy would indicate that the Act is seeking to:

- implement practical measures to protect Australians against exposure to illegal and harmful online content, with particular regard to the needs of Australian children;
- articulate clear expectations of the online services sector as to its responsibilities to keep Australians safe online;
- require appropriate accountability, transparency and user safeguards from online services sector;
- provide a safety net for users where the online services sector fails to meet its obligations under the Act;
- provide a responsive and flexible approach to online safety;
- balance the competing objectives of user safety and freedom of expression;
- empower and encourage the online services sector to develop solutions for online safety risks as far as possible; and
- encourage the development and use of new technologies and safe products and services.

Questions

1. Are the proposed high level objects appropriate? Are there any additions or alternatives that are warranted?
2. Is the proposed statement of regulatory policy sufficiently broad to address online harms in Australia? Are there aspects of the proposed principles that should be modified or omitted, or are there other principles that should be considered?

Basic online safety expectations

Online safety is a shared responsibility. Online service providers can and should play a role in protecting the community from abusive conduct and harmful content online. Some sectors of the technology sector have made significant advances in improving the safety of their services. Online service providers have also worked closely with the eSafety Commissioner and other regulators and agencies on a voluntary basis to address particular issues, as demonstrated through the collaboration on developing the eSafety Commissioner's voluntary SbD principles and the Taskforce to Combat Terrorist and Extreme Violent Material Online.

The Government supports this engagement and expects large technology firms and digital platforms to be proactive in embedding online safety at the outset.

Current approach

There are few requirements under the current legislative framework for online service providers to implement preventative measures to tackle online harms on their services before they occur. There are 'basic online safety requirements' established in the cyberbullying scheme in the EOSA for social media services. These broadly require that:

1. the terms of use (for the social media service) contain a provision that prohibits end-users from posting cyberbullying material on the service;
2. the service has a complaints scheme where users can request the removal of cyberbullying material; and
3. there be a contact point for the eSafety Commissioner.

However, these requirements are limited in application to the cyberbullying scheme (rather than the broader legislative framework) and only apply to social media services. They are also quite contained in their scope. The 2018 Review recommended that the new Act 'guarantee that the online industry goes beyond simple compliance with minimum safety standards and should establish a much higher new benchmark standard with which all industry must comply.'[35]

Without a new set of expectations around minimum standards for pre-emptive and preventative action, there is a risk user safety measures will continue to be reactive, and the burden of safety will continue to fall disproportionately on the end-user.

Safety by Design and the Online Safety Charter

During 2018, the eSafety Commissioner undertook extensive consultation to support the development of a set of voluntary SbD principles. This project identified, in collaboration with the online services sector, a set of voluntary principles to place safety and the rights of users at the centre of the design, development and deployment of online products and services. The next stage of this process will be continued work with the sector to develop specific guidance to assist in the implementation of the principles and a framework to assist companies to embed the principles in their technological development processes.[36]

Building on this work, the Government has released its Online Safety Charter.[37] This is the Government's articulation of community expectations of technology firms and digital platforms to protect citizens, especially children and vulnerable members of the community, from harmful online experiences.

[35] L. Briggs op cit p.2

[36] https://www.esafety.gov.au/key-issues/safety-by-design

[37] Available at www.communications.gov.au

The Charter incorporates these principles as well as some additional areas of focus. The Charter is relevant to technology firms that offer users in Australia the opportunity to interact or connect, and technology firms whose services and products enable users to access content and information. This includes social media services, internet service providers, search engine providers, content hosts, app developers, and gaming providers, among others.

SbD and the Online Safety Charter are complementary:

- SbD is an ongoing consultative process between the eSafety Commissioner and the online services sector to guide systemic shifts in the design of services.
- The Charter is a statement of online safety expectations from Government, citing SbD principles as best practice.

The development of guidance for online service providers on the implementation of SbD will serve to assist the sector to meet the requirements in the Charter and the new basic online safety expectations proposed below.

Proposals

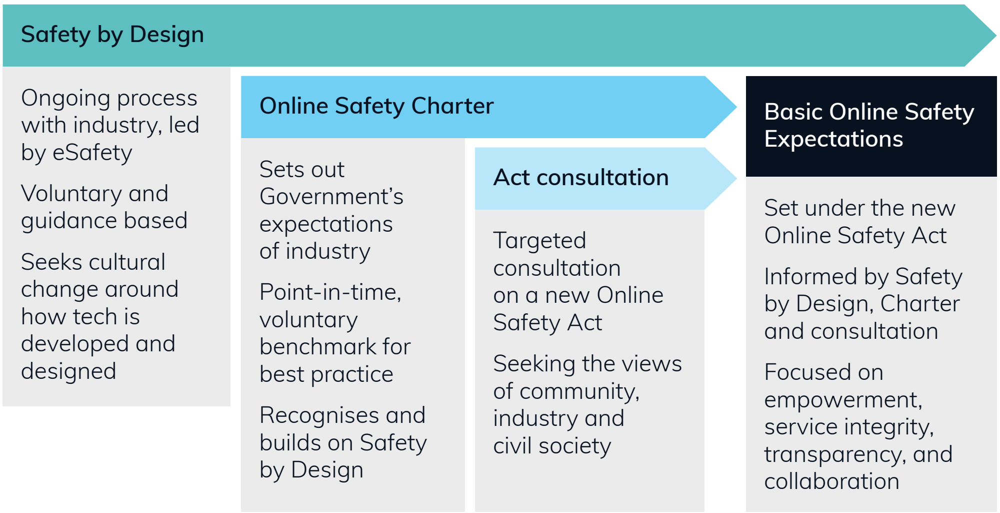
Expansion of the existing obligations

A new Online Safety Act would provide a power for the Minister, via a disallowable legislative instrument, to articulate a set of basic online safety expectations (BOSE). These expectations would be informed by SbD principles, the Online Safety Charter, priorities outlined in the Government's 2019 federal election commitments, and feedback on this discussion paper. The relationship between each of these elements is shown in Figure 3. It is expected the expectations will focus on:

- **empowerment** of users (through effective terms of service, complaint and reporting mechanisms and providing online safety information and advice);
- **transparency** of online service providers' commitment to online safety;
- upholding the **integrity of services** (through effective enforcement of terms of use and proactive measures to identify and remove harmful content); and
- **collaboration** with government and civil society.

Comments are sought through this consultation process on whether the Online Safety Charter is a sufficient basis for the BOSE, or if other additional matters should be addressed.

Figure 3: Contribution of Safety by Design, Online Safety Charter and consultation to the Basic Online Safety Expectations



Safety by Design

Ongoing process with industry, led by eSafety

Voluntary and guidance based

Seeks cultural change around how tech is developed and designed

Online Safety Charter

Sets out Government's expectations of industry

Point-in-time, voluntary benchmark for best practice

Recognises and builds on Safety by Design

Act consultation

Targeted consultation on a new Online Safety Act

Seeking the views of community, industry and civil society

Basic Online Safety Expectations

Set under the new Online Safety Act

Informed by Safety by Design, Charter and consultation

Focused on empowerment, service integrity, transparency, and collaboration

Applicability

As a starting point, the basic online safety expectations would be applicable to all social media services, rather than just the participants in the two tiers of the current cyberbullying scheme. However, to provide flexibility over time, the eSafety Commissioner would have a power to determine by legislative instrument that the expectations apply to other specified types of service providers based on similar criteria to that required under the transparency reporting criteria, including numbers of reports received and response times to requests.

Transparency reporting

The Government is not proposing to impose sanctions for non-compliance with the proposed basic online safety expectations at this stage, though reserves the right to explore this option in future if expectations are not being met. However, the eSafety Commissioner would have a power to determine, by legislative instrument, that particular entities report on their actions in upholding the expectations, through public reporting and/or reporting on specific items to the eSafety Commissioner. Reporting companies would be determined by the eSafety Commissioner based on a range of criteria that are likely to include the numbers of complaints received by the eSafety Commissioner with regard to specific individual companies. It is expected this would be required of larger social media companies and social media companies on which harms have been found to have been occurring by the eSafety Commissioner.

To support this reporting framework, the eSafety Commissioner would be able to impose penalties for non-compliance with the proposed reporting requirements. This will include the capacity for the eSafety Commissioner to publish a statement that a reporting social media service is not complying with the basic online safety expectations.

To minimise the burden on social media services, a single reporting framework would be established. This would, to the fullest extent possible, integrate the reporting requirements of the proposed basic online safety expectations, the transparency recommendation of the Taskforce to Combat Terrorist and Extreme Violent Material Online, the OECD's voluntary transparency reporting protocol (when completed), and the UK's draft transparency reporting template, developed as part of the UK Government's Online Harms White Paper process. It is not expected that companies would have multiple separate transparency reporting obligations, as this would be duplicative and onerous.

This measure will improve transparency around the number of and type of responses to reports and complaints about illegal abusive and predatory content on platforms, delivering on one of the Government's 2019 election commitments.

The proposal to require transparency reporting in relation to the basic online safety expectations is intended to align with and enhance voluntary industry efforts to improve online safety for all Australians. The Government will assess the response of industry to the transparency mechanisms outlined above. In the event that these processes are found to be inadequate, or if there is insufficient effort to meet the expectations, the Government reserves the right to consider additional regulatory action to require compliance.

Protection of children online

The Government has committed to working with online service providers to make online apps, games and services marketed to children default to the most restrictive privacy and safety settings at initial use or set-up. While many services currently provide the option of privacy and safety settings (for example Microsoft Family or Apple operating systems), information on what is available for consumers is not always transparent and accessible.

The Government is looking for industry to ensure that products marketed to children default to the highest level of privacy and safety at the outset, and to enable consumers to set and adjust these controls as they wish. It would be preferable to have these enhanced safety features developed and implemented voluntarily through an industry wide commitment to safety, consistent with the SbD principles and basic online safety expectations. However, in the event that a sector of the industry or particular service providers don't adopt this as a standard practice, the Government will consider the merits of empowering the eSafety Commissioner to specify, by legislative instrument, that particular types of service, or individual service providers with services marketed to children, default to the most restrictive privacy and safety settings.

Point of purchase information

It is important to raise awareness of the tools that are available to parents and carers to protect children online. The eSafety Commissioner has developed many useful resources about appropriate measures for different age groups including how to install parental controls. However, these may not be front of mind when installing a new app, signing up for an account for a child, or purchasing a device at a physical or online store.

While some online service providers already make information available to consumers about online safety and options for parental control of some products and services, these are not always accessible or consistently applied, making it challenging for consumers to make informed decisions about products and services.

The Government has committed to improving the availability of this type of information at all points in the supply chain. To support this aim, it is proposed that the eSafety Commissioner promote the availability of its resources to online and physical stores and could, subject to resourcing considerations, provide copies of materials and guides. However, should particular sectors or particular retailers opt not to provide this type of information, the Government will consider the merits of empowering the eSafety Commissioner to make rules, by legislative instrument, requiring point of sale information on online safety features and parental control settings.

Questions

3. Is there merit in the BOSE concept?
4. Are there matters (other than those canvassed in the Charter) that should be considered for the BOSE? Are there any matters in the Charter that should not be part of the BOSE?
5. What factors should be considered by the eSafety Commissioner in determining particular entities that are required to adhere to transparency reporting requirements (e.g. size, number of Australian users, history of upheld complaints)?
6. Should there be sanctions for companies that fail to meet the BOSE, beyond the proposed reporting and publication arrangements?

Cyberbullying scheme

Cyberbullying poses a risk to children

According to research conducted by the eSafety Commissioner, 1 in 5 children have experienced cyberbullying. The top five negative online experiences of young people between 8 and 17 years of age, and the percentage of respondents experiencing them, were:

Being contacted by strangers/someone they did not know	**25%**
Being left out by others	**21%**
Having mean things said about them/called names	**19%**
Receiving repeated unwanted online messages from someone	**13%**
Having lies/rumours spread about them	**13%**[38]

Young people aged 8 to 17 are often unwilling to or unaware that they can seek help from parents, carers, educators, digital platforms and authorities. A study conducted in 2017 found that although 71 per cent of young people who had negative online experiences sought help in an informal capacity through families and friends, only 24 per cent sought help in a formal way.[39]

A December 2018 survey by mental health service provider ReachOut of 1000 young people aged 14 to 25 found that 24 per cent of them had been bullied and that for 36 per cent of these respondents, the bullying had been online (8.7 per cent of the total). The survey also found that 15 per cent of respondents dealt with bullying by using drugs or alcohol.[40]

Parents and carers are concerned about online safety risks for children. The three most common concerns cited in research undertaken by the eSafety Commissioner in 2018 were: exposure to inappropriate content other than pornography (38 per cent); contact with strangers (37 per cent); and being bullied online (34 per cent).[41]

Current approach

The EOSA currently sets out a two-tiered complaints scheme for the rapid removal of cyberbullying material targeted at an Australian child. The scheme has been successful in providing victims of cyberbullying with a simple and straight forward means of having cyberbullying material removed from social media services and

[38] eSafety Commissioner, *State of Play – Youth, kids and digital dangers* (2018), p. 21.

[39] Ibid, p. 24.

[40] 'New research finds 16 percent of young people use drugs and alcohol to help them cope with bullying' 4 March 2019 ,https://about.au.reachout.com/bullying-2019

[41] https://www.esafety.gov.au/-/media/cesc/esafety-corporate/research/esafetyresearchparentingdigitalage.pdf

relevant electronic services. This success is the product of the development of an effective and collaborative partnership with online service providers, and positive and proactive engagement from social media providers.

- Since July 2015, the eSafety Commissioner has received more than 1,500 complaints regarding the cyberbullying of young Australians, and has had a 100 per cent success rate in seeking the removal of cyberbullying material from the 'tier partner' social media services (outlined below).
- The take-down of cyberbullying material has been prompt and in certain cases as quick as within 30 minutes from the request to its removal from the service.

Under the scheme, there are two tiers of regulation that apply to social media services. These are outlined in Table 1.

Table 1: Social media services regulated under the cyberbullying scheme

Tiers of social media services	Relevant social media services
Tier 1 services – social media services that have applied to and satisfied the eSafety Commissioner that they can comply with the basic online safety expectations. Can be asked by the eSafety Commissioner to remove cyberbullying material targeted at an Australian child.	airG, Ask.fm, Flickr, TikTok (formerly Musical.ly), Roblox, Snapchat, Twitter, Yahoo!7 Answers, Yahoo!7 Groups, and Yubo.
Tier 2 services – social media services that have been declared as Tier 2 by the Minister, following a recommendation by the eSafety Commissioner. Can be required by the eSafety Commissioner to remove cyberbullying material targeted at an Australian child.	Facebook, Instagram and YouTube[42]

[42] Google+ is also a Tier 2 service but is no longer being provided by Google.

Material is considered cyberbullying if it satisfies the following conditions:

- The material is provided on a social media service or relevant electronic service.
- An ordinary reasonable person would conclude that:
  - it is likely that the material was intended to have an effect on a particular Australian child; and
  - the material would be likely to have the effect on the Australian child of seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child.
- The material satisfies other conditions set out in relevant legislative rules (of which there are none at this time).

A complaint may be made to the eSafety Commissioner when a person has reason to believe cyberbullying material targeted at an Australian child has been, or is being provided on a social media service or relevant electronic service. The person must demonstrate that they have, in the first instance, made a complaint to the social media service under its existing complaints system.

If the material was not removed from the service within 48 hours of the initial complaint to the service and the eSafety Commissioner is satisfied that the material was cyberbullying material directed at an Australian child, they may:

- For a Tier 1 service, request that the material be removed within 48 hours.
- For a Tier 2 service, require that the material be removed within 48 hours

For the purposes of the EOSA, material is *removed* if the material is neither accessible to, nor delivered to, any of the end-users in Australia using the service.

Following the removal of any material through the eSafety Commissioner's complaints scheme, if material is reposted, a further request for take-down is required to be made to the eSafety Commissioner.

The eSafety Commissioner may also issue end-user notices that require a person who posts cyberbullying material to remove the material, refrain from posting any cyberbullying material targeting the child, and/or apologise to the child for posting the material.

The eSafety Commissioner has in place processes and an agreement with the Australian Federal Police so that, where required, material is retained for law enforcement purposes, and that appropriate referrals are in place in the case of criminal conduct.

While the cyberbullying scheme has proven to be an effective mechanism for addressing cyberbullying material directed at an Australian child, there are challenges as harmful conduct occurs on new services and platforms.

- Messaging applications, such as WhatsApp, Signal and Telegram, are increasingly popular due to their ability to securely exchange text, voice and images within private groups. Seventy five per cent of Australian internet users had used an app to communicate via messages, voice or video calls in the six months to May 2018.[43] There are mechanisms available to address cyberbullying directed at an individual on these services, as users generally have the ability to block unwanted communication, delete abusive messages, export chat logs and report users to service provider. However, the capacity for victims to effectively address cyberbullying conduct occurring within message groups – where there is typically a large number of recipients – is more limited.
- Australian children are also keen users of online gaming platforms, many of which provide in-game communication systems that can potentially be used to bully or abuse other players. Bullying within online games creates content that is often ephemeral (only available for a time-limited period) and cannot be easily reported to the service provider or to the eSafety Commissioner.

Proposals

Broadening ranges of service providers covered

The current two-tier arrangements mean that, in effect, mandatory removal notices for cyberbullying material directed at an Australian child can only be issued to large social media services.[44]

This construction no longer aligns with the patterns of cyberbullying occurring in Australia and overseas. It is clear, from research and the experience of the eSafety Commissioner in administering the current scheme, that cyberbullying is no longer occurring on a single platform, and certainly not just on the larger social media services.[45,46] Bullying, abuse and harassment occurs across a range of platforms and services of various sizes and types including:

- gaming, game streaming and game chat services (like Twitch, Fortnite and Discord);
- messaging apps (like WhatsApp, Kik, WeChat, Viber, GroupMe, Jott and Tango);
- 'confessional' platforms (like Tellonym and Whisper); and
- social connection sites (like Yubo, Holla, MeetMe and Monkey).

[43] Australian Communication and Media Authority, *Communications Report* 2017–18

[44] Tier 1 social media services may be requested to remove the material under a voluntary basis, while Tier 2 social media services may be given a notice *requiring* the removal of material.

[45] https://www.ofcom.org.uk/__data/assets/pdf_file/0028/149068/online-harms-chart-pack.pdf

[46] https://www.pewinternet.org/wp-content/uploads/sites/9/2018/05/PI_2018.05.31_TeensTech_FINAL.pdf

Not all of these platforms are currently members of Tier 1 or Tier 2 as they have not been declared by the Minister to be a Tier 2 service, or because they have not approached the eSafety Commissioner to demonstrate they can comply with the current basic online safety requirements under Tier 1.

In the interests of harmonising and streamlining the coverage of the online safety schemes, it is proposed that a new Online Safety Act would expand the cyberbullying scheme to cover the same categories of service included in the image-based abuse scheme:

- social media services;
- relevant electronic services; and
- designated internet services.

Consistent with the image-based abuse scheme, it would also include services that *host* a social media service, a relevant electronic service or a designated internet service.

Additional tools for the eSafety Commissioner to address cyberbullying

There may be merit in providing the eSafety Commissioner with additional tools to address cyberbullying of Australian children. These would operate in addition to the existing social media service or end-user notice arrangements, and would be designed to be used where social media service or end-user notices are not well suited to the particular type of content or service on which the cyberbullying has taken place.

These types of additional tools could include the capacity for the eSafety Commissioner to either request, or require, that a platform or service provider enforce their terms of service in relation to a user who has been found to have posted cyberbullying material, apply account restrictions in serious cases, or to request or require certain other enforcement actions. These types of alternative tools may be particularly useful in dealing with cyberbullying occurring on services such as gaming platforms or messaging applications.

Shortening the response time from 48 to 24 hours

In recognition of how harmful material can be if it remains online, and the need to act quickly, it is proposed that the compliance time for online service providers responding to the eSafety Commissioner's removal notices be shortened from 48 to 24 hours.

This is based on the experience of the eSafety Commissioner and the prompt removal times online service providers have shown they are capable of achieving on a voluntary basis. It is also consistent with international practice for take-down of illegal and harmful content.

- In Germany, platforms are required to remove illegal content in 24 hours.[47]
- The French government has introduced legislation to require platforms to remove overly hateful content with 24 hours.[48,49]

Questions

7. Is the proposed expansion of the cyberbullying scheme for children to designated internet services and hosting services, in addition to relevant electronic service and social media services, appropriate?
8. Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?
9. What are the likely compliance burdens of the proposed changes to the cyberbullying scheme on small and large businesses?
10. What other tools could the eSafety Commissioner utilise to effectively address cyberbullying in the circumstances where social media service and end-user notices are not well suited to the particular service upon which the cyberbullying has occurred?

Establishing a new cyber abuse scheme for adults

Online harms also impact adults

While it is globally recognised that children are particularly vulnerable to online harms, adults in the community can also be vulnerable, for a range of different reasons. These can include, but are not limited to, resilience and mental health, structural, socio-economic and environmental reasons, risk-taking behaviours, exposure and scale, and social, regulatory and technological contexts.

When first introduced, the eSafety Commissioner's work focused on protecting children. This remit was expanded in 2017 to include all Australians, recognising the importance of online safety for the community at large and the changes occurring in digital technology and user behaviour. However, this remit did not extend to all of the schemes overseen by the eSafety Commissioner. In particular, the cyberbullying reporting and take-down scheme continues to be only available to children.

[47] https://transparencyreport.google.com/netzdg/youtube

[48] https://www.theguardian.com/world/2019/jul/09/france-online-hate-speech-law-social-media

[49] http://www.assemblee-nationale.fr/15/propositions/pion1785.asp

Recent statistics on the prevalence of online harms for adults demonstrate the need to address this gap:

- In 2018, Amnesty International undertook a poll in Australia on the experiences of women aged between the ages of 18 and 55 and found that three in ten women surveyed had experienced online abuse or harassment, and nearly half for respondents aged 18 to 24, with 37 per cent saying the experience had made them feel physically unsafe.[50]
- A national study undertaken by Ofcom in the UK similarly found that three in ten adult internet users had experienced something they rated as harmful over the last twelve month period.[51]
- Plan International's 2019 snapshot of social media commentary of sportswomen and sportsmen found that 'more than a quarter of all comments towards sportswomen were sexist, sexualised, belittled women's sports or were otherwise negative in nature'.[52]
- The 2015 United Nations *Cyber Violence against Women and Girls* report suggested that 73 per cent of women worldwide had experienced some form of online violence.[53]

Current approach

The *Commonwealth Criminal Code Act 1995* already includes criminal offences for using a carriage service to menace, harass or cause offence (Section 474.17). This section has been used to prosecute serious cyber abuse of adults. However, there is currently no equivalent of the cyberbullying takedown regime for material aimed at an adult.

The 2018 Review of online safety legislation recommended the eSafety Commissioner's remit be extended to cover all adults experiencing cyberbullying. The Review also noted the tight limitation of the eSafety Commissioner's role with respect to adults contradicts the experience of many Australians with regard to online harassment and trolling, especially women with high profiles, Aboriginal and Torres Strait Islander women, Islamic spokespeople, and the families of murder and rape victims.[54]

[50] https://www.amnesty.org.au/australia-poll-reveals-alarming-impact-online-abuse-women/

[51] https://www.ofcom.org.uk/research-and-data/internet-and-on-demand-research/internet-use-and-attitudes/internet-users-experience-of-harm-online-2019

[52] https://www.plan.org.au/learn/who-we-are/blog/2019/04/24/240419-snapshot-analysis

[53] https://www.unwomen.org/~/media/headquarters/attachments/sections/library/publications/2015/cyber_violence_gender%20report.pdf?v=1&d=20150924T154259&v=1&d=20150924T154259

[54] L. Briggs, op cit , p 33

Proposal

It is proposed the new Online Safety Act include a new scheme for cyber abuse of adults. The focus of the new cyber abuse scheme would be on serious cases of abuse, recognising that adults can be expected to demonstrate a higher level of resilience and maturity than children, and that it will be important to avoid creating an unreasonable regulatory burden. The cyber abuse scheme would aim to provide a safeguard for serious instances of online harassment and humiliation, cyberstalking, including threats to cause harm, or online abuse experienced as part of domestic and family violence.

This proposal is about creating a pathway for the take-down of material, focussing on minimising harm to the victim of the abuse. For the purposes of this consultation process, a proposed definition of adult cyber abuse material is provided below.

> For online material to meet the statutory test of being cyber abuse material targeting an Australian adult, it would need to satisfy all of the following conditions:
>
> - the material would be provided on a social media service, relevant electronic service, or designated internet service;
> - an ordinary reasonable person would conclude that the material was intended to have an effect of causing serious distress or serious harm to a particular Australian adult; and
> - an ordinary reasonable person would, in all the circumstances, regard the material as menacing, harassing or offensive, whether because of the manner in which the material is provided, or the content of the material, or both.

This definition is intended to set a higher threshold for what constitutes adult cyber abuse compared with the cyberbullying of an Australian child.

- For adults, the material in question would need to be intended to have an effect of causing **serious distress or harm**, rather than intended to have **an effect on the person**.
- For adults, the material would need to be **menacing, harassing or offensive** (taking into account all of the circumstances), mirroring the construction of offence provisions under section 474.17 of the *Criminal Code Act 1995*, rather than **likely to have the effect of** seriously threatening, intimidating, harassing or humiliating.

The definition of 'cyberbullying material targeted at an Australian child' in the Act would not be altered. The current cyberbullying definition is working effectively and the scheme provides the eSafety Commissioner with sufficient scope and flexibility to address bullying directed at children.

As noted in relation to the current cyberbullying scheme, the current social media service and end-user notices available to the eSafety Commissioner to address cyberbullying content under the current scheme may not be well suited to the full range of services on which cyberbullying is occurring, including gaming platforms and messaging apps. As with the proposed reforms to the cyberbullying scheme, comment is sought on the additional tools that could be made available to the eSafety Commissioner under the cyber abuse scheme. These could include the capacity for the eSafety Commissioner to either request, or require, that a platform or service provider enforce their terms of service in relation to a user who has been found to have posted cyber abuse material, apply account restrictions in serious cases, or to request or require certain other enforcement actions.

Civil penalties

The non-consensual sharing of intimate images scheme currently prohibits a person from, or making a threat to post, an intimate image without consent (section 44B of the EOSA). Doing so attracts a civil penalty of up to $105,000. A person must comply with the requirement under a removal notice to the extent they are capable of doing so, or they may face a civil penalty (section 44G of the EOSA) (see **Attachment A**).

It is proposed that the establishment of a cyber abuse scheme for Australian adults would include an equivalent end-user take-down and penalty regime. It is not proposed to extend this penalty framework to the cyberbullying scheme for children.

This civil penalty scheme would not override or supplant existing criminal provisions for abuse and harassment. Under the *Criminal Code Act 1995*, and at the state and territory level, end users can be subject to prosecution relating to stalking, threats to kill or cause serious harm, making hoax threats, engaging in conduct that a reasonable person would find to be menacing, harassing or offensive and promoting suicide, among others.

Separately, the Government has committed to strengthening existing criminal penalties for online abuse and harassment, and developing national principles and a consistent approach to combatting criminal cyberbullying and online harassment, including seeking to address inconsistencies between approaches to criminal cyberbullying across Australia. This work is underway and will be an ongoing collaboration across jurisdictions.

The eSafety Commissioner would continue to work closely with law enforcement to enable the referral of material to relevant law enforcement where this is warranted. Any reform to the civil penalties scheme would also need to consider principles of youth justice, in consultation with law enforcement agencies.

Questions

11. Is the proposed application of the cyberbullying and cyber abuse schemes to designated internet services and hosting services, relevant electronic service and social media services, appropriate?
12. Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?
13. Do the proposed elements of a definition of adult cyber abuse appropriately balance the protection from harms with the expectation that adults should be able to express views freely, including robust differences of opinion?
14. Should the penalties differ under a cyber abuse scheme for adults and the cyberbullying scheme for children?
15. What additional tools or processes, in addition to removal notices, could be made available to the eSafety Commissioner to address cyber abuse occurring across the full range of services used by Australians?

Non-consensual sharing of intimate images (image-based abuse)

Image-based abuse affects a range of groups

In 2017, image-based abuse, or the sharing of intimate images without consent, was found to have affected 1 in 10 Australians. This is an extremely destructive form of online abuse which can have devastating impacts for victims.[55]

The sharing of intimate images without consent is, at times, linked to intimate partner and family violence situations, with 1 in 4 women reporting have experienced emotional abuse from a former or current partner, and 1 in 6 reporting having experienced physical violence in 2016.[56] According to 2017 research by the eSafety Commissioner, image-based abuse is more prevalent amongst certain population groups including Australians from Aboriginal or Torres Strait Islander descent (25 per cent), younger women (24 per cent) and those who identify as LGBTI (19 per cent).[57]

The eSafety Commissioner has also reported that in 98 per cent of domestic and family violence situations, technology-facilitated abuse is an extension of this realworld violence where victims are abused and stalked through the use of technology.[58]

[55] https://www.esafety.gov.au/sites/default/files/2019-07/Image-based-abuse-national-survey-summary-report-2017.pdf

[56] Australian Institute of Health and Welfare, *Family, domestic and sexual violence in Australia: continuing the national story* (2019)

[57] eSafety Commissioner, *Image-Based Abuse National Survey: Summary Report*, October 2017

[58] https://www.esafety.gov.au/about-the-office/newsroom/blog/a-big-year-keeping-australians-safer-online

These findings are consistent with the results of the 2016 RMIT research of more than 4,000 participants about the sharing of intimate images. This found that 23 per cent of respondents had been subject to some form of image-based abuse and that men and women were equally likely to report being a victim. Members of more vulnerable groups in the community reported a higher incidence of abuse, with one in two respondents with a disability and one in two Indigenous respondents. More than one in three respondents in the LGTBIQ+ community also reported being subject to abuse.[59]

Current approach

The image-based abuse scheme came into effect with the passage of the *Enhancing Online Safety (Non-consensual Sharing of Intimate Images) Act 2018*. The scheme applies to end-users, social media services, designated internet services, relevant electronic services and hosting services.

The eSafety Commissioner may issue removal notices that require the providers of social media services, relevant electronic services, designated internet services and hosting services to take all reasonable steps to support the removal of intimate images, or to cease hosting the image. The eSafety Commissioner may also issue a removal notice to the person posting an image (an end-user notice).

The eSafety Commissioner may issue informal or formal warnings in relation to the contravention of the prohibition on posting an intimate image, the failure to ensure the removal of an intimate image or a contravention of a remedial direction (in relation to the posting of an intimate image). Other enforcement options available to the eSafety Commissioner include infringement notices, enforceable undertakings and injunctions. The penalties available under this scheme, and the other existing online safety schemes, are outlined at **Attachment A**.

Between the start of the civil penalties scheme on 1 September 2018 and 30 June 2019, the eSafety Commissioner:

- received 849 reports of image-based abuse;
- issued one removal notice and three formal warnings to persons responsible for image-based abuse; and
- issued eight informal warnings to persons responsible for image-based abuse, adopting an educative approach to enforcement in appropriate cases.[60]

Where image-based abuse has been found to have occurred on services such as messaging apps, the eSafety Commissioner has, in some cases, opted to advise service providers of the abuse, and asked them to take appropriate action under their relevant terms of service.

[59] RMIT research - N.Henry, A. Powell, A. Flynn ‘Not Just ‘Revenge Pornography’: Australians’ Experiences of Image-Based Abuse A SUMMARY REPORT May 2017.

[60] eSafety Commissioner *Office of the eSafety Commissioner Annual Report 2018–19*, p. 208

Overall, the eSafety Commissioner has been successful in having image-based abuse material removed in more than 80 per cent of cases, despite nearly all websites reported to date being hosted overseas.

Proposals

It is not proposed to substantively change the operation of the image-based abuse scheme. The scheme is modern, has appropriate coverage of services, and is operating effectively.

However, in recognition of how harmful this type of material can be to the victim if it remains online, the timeframe for online service providers to comply with the eSafety Commissioner-issued removal notices would be shortened from 48 to 24 hours. This is based on international practice, reflects the prompt removal times the eSafety Commissioner has achieved to date, and would align the take-down timeframes across all four proposed online safety schemes.

In addition, the definition of 'intimate images' would be amended to cover intimate images which purport to be a person. This is not clear from the current wording in the EOSA and is an issue that has emerged over recent years.

Consistent with the cyberbullying and cyber abuse schemes, comment is sought on the merits of the eSafety Commissioner being provided with additional powers to address image-based abuse, beyond the removal notices already available.

Questions

16. Is the proposed take-down period for the image-based abuse scheme of 24 hours reasonable, or should this require take-down in a shorter period of time?
17. Does the image-based abuse scheme require any other modifications or updates to remain fit for purpose?
18. What additional tools or processes, in addition to removal notices, could be made available to the eSafety Commissioner to address image-based abuse being perpetrated across the range of services used by Australians?

Addressing illegal and harmful online content

Illegal and harmful content online is an extensive and global problem

The framework currently governing harmful online content is not suited to the contemporary online environment and the technologies and services used by Australians every day. The highest emerging risk is with nonprofessional, user-generated content. Digital platforms often face challenges in moderating the vast volumes of content being uploaded/created by their users even with the help of automation.[61] For example, in 2019, YouTube reportedly had more than 500 hours of video uploaded every minute.[62]

In the 2018-19 financial year, the eSafety Commissioner completed 12,126 statutory investigations into online content. This included more than 8,000 investigations into child abuse content. The removal of these images helps to reduce the risk of survivors being further victimised.[63]

However, this enforcement response addresses only a tiny part of a global problem.

› In 2018, the US-based National Centre for Missing and Exploited Children received 18.4 million reports of online child sexual abuse, containing more than 45 million images and videos.

› Of these reports, 16 million were made by the digital industry (approximately 12 million by Facebook alone).[64]

There is also an international network of online safety hotlines, called INHOPE. This has 48 members in 43 countries including Australia. It works with law enforcement and online service providers in these countries to investigate and remove child sexual abuse material. The eSafety Commissioner refers overseas hosted illegal content to INHOPE for investigation. In 2018, INHOPE reports that in response to 155,240 reports, it identified 223,999 images and videos of child sexual abuse material and was able to have 58 per cent of it removed within three days.[65]

INHOPE does not address the full range of material that affects online safety: just the worst of the worst.

Current approach

The online content scheme seeks to address the publication of illegal and offensive material online ('prohibited and potential prohibited content') and prevent children from being exposed to material that would be likely to offend a reasonable adult.[66] The online content scheme is established by Schedules 5 and 7 of the BSA.

The existing scheme establishes limits on the types of online content that can be provided or hosted by internet service providers and content service providers, respectively, and provides mechanisms for users to complain to online service providers or the eSafety Commissioner about prohibited or potential prohibited content. This is content that may be:

[61] https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#387846d860ba

[62] https://youtube.googleblog.com/2019/09/appealspeech.html

[63] eSafety Commissioner, *Office of the eSafety Commissioner Annual Report 2017–18*, p.127

[64] https://www.ecpatusa.org/blog/tag/child+sexual+abuse+imagery

[65] INHOPE 2018 Statistics Report, http://88.208.218.79/Libraries/IC-CAM_IHRMS/INHOPE_Statistics_Report_2018.sflb.ashx

[66] https://www.legislation.gov.au/Details/C2007B00081/Explanatory%20Memorandum/Text

- Refused Classification, or RC, under the National Classification Code, which includes:[67]
  - illegal material such as child sexual abuse material;
  - extremely violent and disturbing pornography;
  - extremist propaganda, incitement to terrorism; and
  - games that victimise and abuse children or encourage illegal activity; and
- X18+ content that contains real depictions of actual sexual activity between consenting adults without violence, coercion or other types of abuse.

The online content scheme also seeks to restrict access by children to content that may be suitable for adults, but not children, including:

- R18+ content which may for example contain violence, drug use, nudity or realistically simulated sex; and
- MA15+ content on certain mobile premium services, or that is commercially provided (other than text and/or still images).

This framework operates as a co-regulatory system is supported by industry codes. Under these industry codes, commercial content providers and certain mobile content services assess some content in advance of uploading, and assess uploaded content in response to complaints, and then apply the appropriate measures to manage end-users' access, which may involve take-down (including link and service deletion), blocking technology to prevent distribution, or access controls, such as restricted access systems like PINs and credit card age verification. The codes also require industry to respond to notices and help parents monitor the online activities of their children and filter unwanted content.

The eSafety Commissioner investigates complaints about prohibited or potential prohibited content. If the content is hosted in Australia, the eSafety Commissioner can order the take-down of material using powers in Schedule 7. The eSafety Commissioner can also require live-streamed content services and services that provide links to content to take certain remedial actions. If the content is hosted outside of Australia, the eSafety Commissioner must report it to law enforcement if it is of a sufficiently serious nature and advise links to the makers of internet filters using powers under Schedule 5.

[67] Clause 20 and 21 of schedule 7 of the BSA refers to the classifications in the National Classification Code 2005. The code is developed by Commonwealth, state and territory governments to reflect community standards. The classification codes are used to classify film, games and publications.

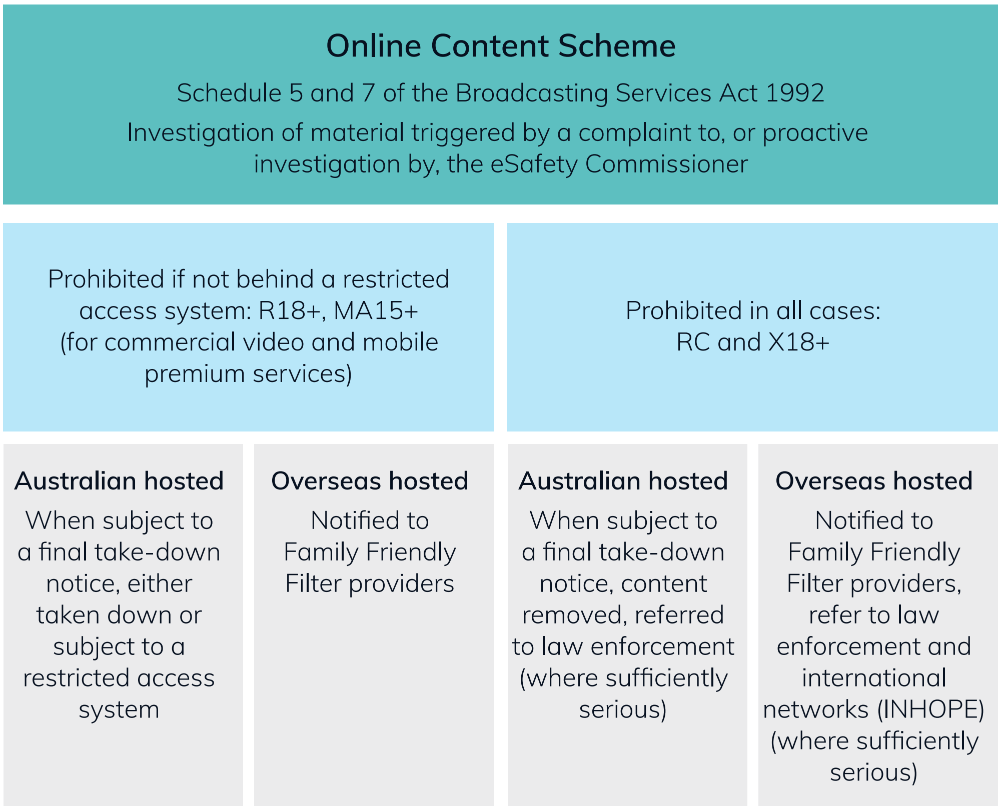
The case for reform

Content regulation practices developed for traditional broadcast, film and print publications are not suited to the volumes of content that are now uploaded each day. Schedules 5 and 7 to the BSA, and the four industry codes made under them, were developed in a pre-smartphone/pre-social media world and no longer reflect the reality of the online experiences, digital technologies and consumption of content by Australian consumers today.

The eSafety Commissioner has highlighted that the current framework is misaligned with current technologies, usage patterns, community concerns and enforcement mechanisms, especially in relation to some of the most concerning content.

Schedules 5 and 7 contain multiple provisions relating to enforcement powers with escalating severity for non-compliance. According to the eSafety Commissioner, most of the enforcement powers have not been used since the eSafety Commissioner was established.

Figure 4: Summary of current Online Content Scheme



Online Content Scheme

Schedule 5 and 7 of the Broadcasting Services Act 1992

Investigation of material triggered by a complaint to, or proactive investigation by, the eSafety Commissioner

Prohibited if not behind a restricted access system: R18+, MA15+ (for commercial video and mobile premium services)

Prohibited in all cases: RC and X18+

Australian hosted

When subject to a final take-down notice, either taken down or subject to a restricted access system

Overseas hosted

Notified to Family Friendly Filter providers

Australian hosted

When subject to a final take-down notice, content removed, referred to law enforcement (where sufficiently serious)

Overseas hosted

Notified to Family Friendly Filter providers, refer to law enforcement and international networks (INHOPE) (where sufficiently serious)

Proposals

The restrictions on harmful online content (RC, X18+ content, R18+ and MA15+ content) hosted by Australian online service providers will remain, in accordance with community expectations. However, the way that these types of content will be addressed will be updated, with a stronger role for industry to prevent exposure to harmful content and provide consumers with greater control over the types of content surfaced on relevant services.

Principles-based codes for industry to address harmful content

The majority of submissions to the 2018 Review of online safety legislation proposed taking schedules 5 and 7 out of the BSA and incorporating key elements within the new Online Safety Act. The migrated elements would retain the provisions for online service providers to develop codes of practice to address harmful online content and for the eSafety Commissioner to make an industry standard should these codes prove to be ineffective. The eSafety Commissioner would retain powers to refer sufficiently serious content to law enforcement for investigation.

However, the code provisions would be updated to require codes to be principles-based and stipulate that codes should be developed by a wider range of service providers than the current codes, reflecting the range of online services that Australians now use to access online content.

The obligations and actions expected of industry through the codes would take into account the nature and characteristics of particular online services. For example, ISPs do not moderate or publish content, and the expectations for these providers would be tailored to their roles. In contrast, designated internet services, such as websites, have a greater degree of control over the content made available through their services. To this end, it is expected that the codes applicable to these types of services would require that content that would otherwise be classified RC or X18+ must not be hosted in Australia. Breaching a code provision relating to the hosting of X18+ or RC would be treated very seriously and could trigger an investigation by the eSafety Commissioner into the content host.

Despite these differences, a consistent feature of the codes would be the requirement for all sectors of industry to provide their users with access to the best available technology solutions to help Australian families to limit access to prohibited content, whether it is hosted in Australia or offshore. The kinds of tools that could be used under the codes are explored further in the section **Opt-in tools and services to restrict access to inappropriate content**.

The concept of harmful content under the codes would be informed by the National Classification Code, and the technology solutions deployed would be proportionate to the potential harm posed by the material.

The codes would be developed by industry in consultation with stakeholders and would require approval by the eSafety Commissioner before coming into effect.

The codes would establish complaints-handling processes for members of the public to report noncompliance. Investigation of code breaches could, for example, involve a stepped process of submitting an initial complaint to the service, then escalating the matter to the eSafety Commissioner.

The new codes would be backed up by a standard-making power so the eSafety Commissioner could intervene and create binding rules for industry in the event that the codes are not operating effectively. The eSafety Commissioner would continue to have access to a range of options to enforce compliance, including warnings, notices, undertakings, remedial directions and civil penalties.

Strengthen the eSafety Commissioner's ability to address seriously harmful content

Given the success of Australian industry in addressing seriously harmful online content hosted in Australia, the staff of the Office of eSafety Commissioner spend the majority of their time investigating content on overseas hosted sites. The eSafety Commissioner's powers to address seriously harmful content would be strengthened under the new Act. It is proposed that the definition of 'seriously harmful content' be based on content that would be illegal under the Commonwealth Criminal Code including:

- child sexual abuse material;
- abhorrent violent material; and
- content that promotes, incites or instructs in serious crime.

To provide a flexible approach, the Minister, on the basis of advice from the eSafety Commissioner, would be provided with the power to make a legislative instrument to capture additional types of content that meet the threshold for seriously harmful, should they emerge. For example, there may be a new type of material that emerges that is found to be causing harm, such as virtual reality or animated content.

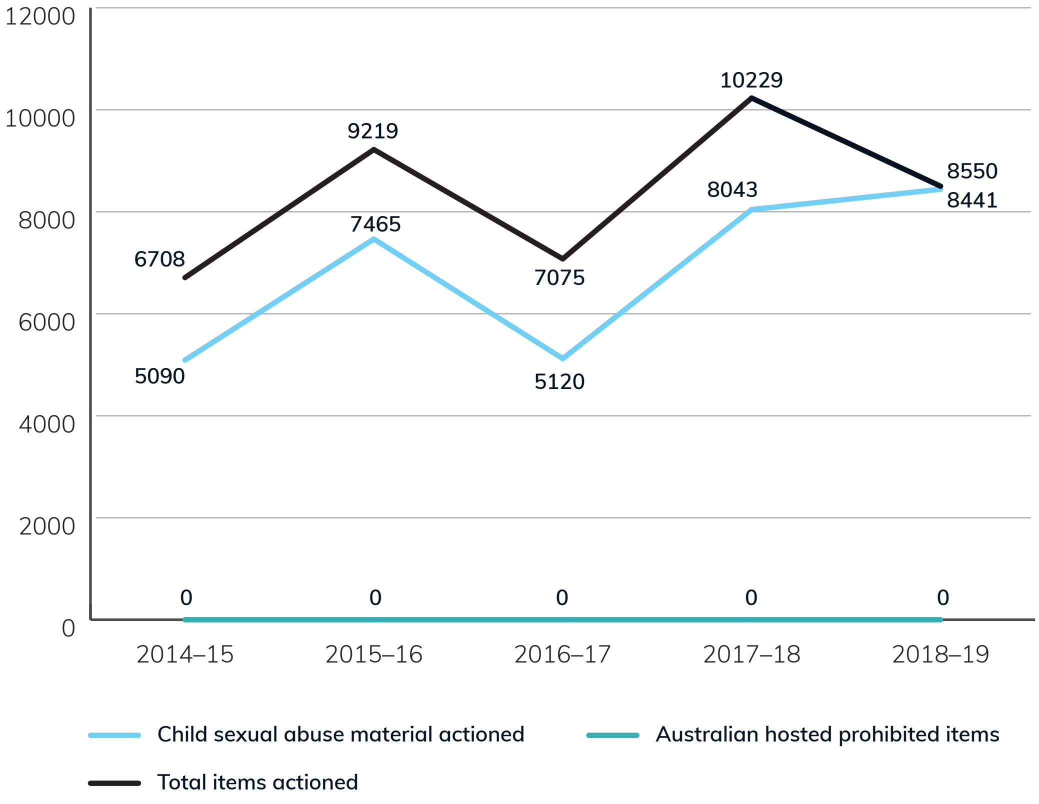
This type of online content would no longer be assessed under the National Classification Code. Instead, the eSafety Commissioner would be able to assess content to determine if it meets the definition of 'seriously harmful content'. In making this determination, the eSafety Commissioner may have regard to the guidance provided by the Classification Code, but the use of the Classification Code and referral to the Classification Board would no longer be mandatory before the eSafety Commissioner could make an assessment of the particular content. Such a process would improve the eSafety Commissioner's ability to respond effectively to this type of content.

Under the current scheme, the eSafety Commissioner is only able to issue take-down notices to content hosted in Australia. It is proposed that this restriction be removed, providing the eSafety Commissioner with the ability to issue take-down notices to social media services, designated internet services and relevant electronic services that provide seriously harmful content that is able to be accessed by Australians, irrespective of whether the content is hosted in Australia or overseas.

This approach would be consistent with the existing take-down notice scheme for image-based abuse and the eSafety Commissioner's powers to issue notices under the AVM Act.

This would also remove the current difference in the action that the eSafety Commissioner can take depending on where content is hosted (take-down for Australian-hosted content, referral to accredited end-user filter providers for overseas-hosted content). This is particularly pertinent given that overseas-hosted material forms the vast majority of the prohibited online content actioned by the eSafety Commissioner (see Figure 6).

Figure 5: Prohibited online content actioned, 2014–15—2018–19



Source: ACMA and eSafety annual reports for 2014-15; 2015–16; 2016–17; 2017–18; 2018–19.

The eSafety Commissioner would continue to be able to refer sufficiently serious content (particularly child sexual abuse material) to relevant law enforcement bodies (i.e. Australian police forces) and other parties (INHOPE network and other similar bodies). This would also be facilitated by empowering the eSafety Commissioner to enter into agreements with other international partners, to facilitate take-down as part of the proposed function of preventing online harms (see below).

It is also proposed to harmonise the take-down timeframes for seriously harmful content with the cyberbullying scheme, proposed cyber abuse scheme and image-based abuse scheme. This would require the removal of seriously harmful content following a notice within 24 hours (current timeframes are to take down as soon as practicable, and in any event by 6:00 pm on the next business day after the notice was given). A 24 hour take-down timeframe would be consistent with the German NetzDG law and the EU Code of Conduct on Countering Illegal Hate Speech Online.

These arrangements to address seriously harmful content would apply to the three main categories of service that are already subject to the image-based abuse scheme:

- social media services;
- designated internet services (including websites); and
- relevant electronic services (such as messaging services, chat services and SMS).

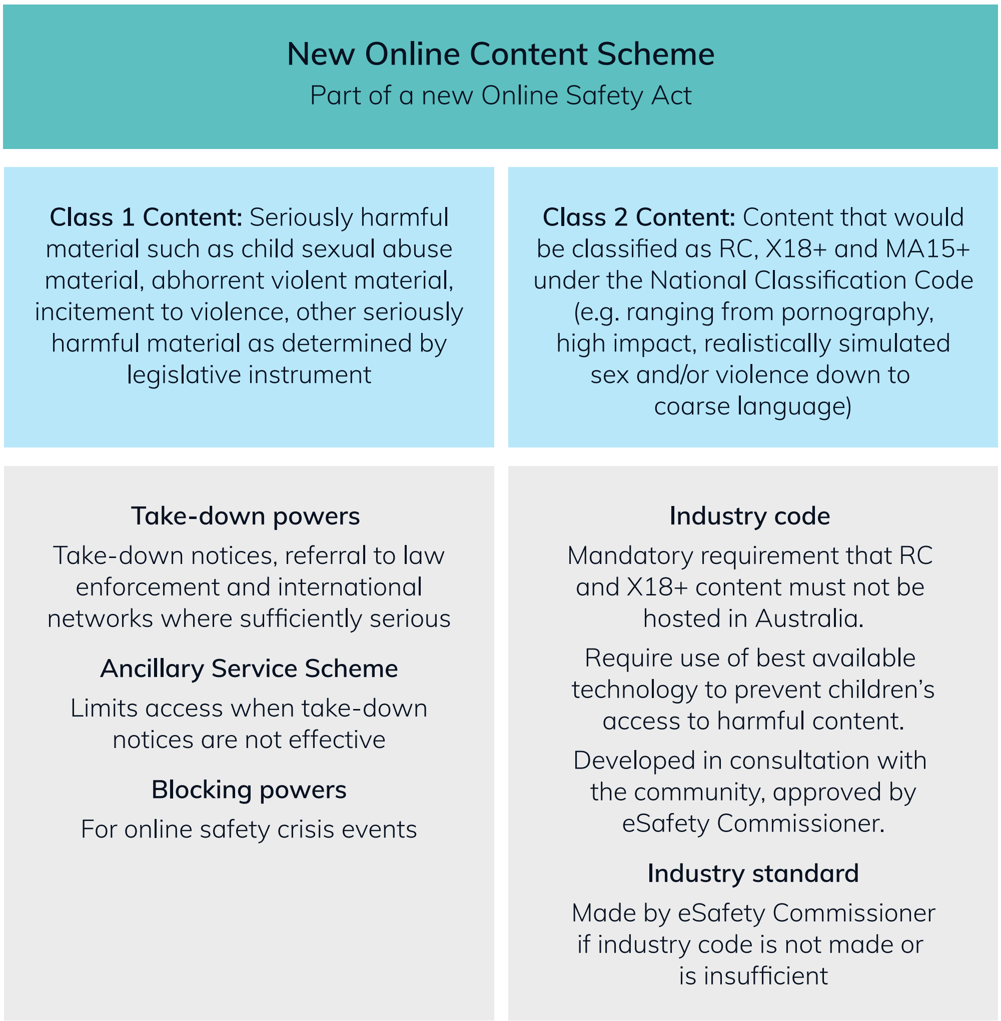
This would create a harmonised set of obligations for the take-down of seriously harmful content across the four schemes (cyberbullying, cyber abuse, image-based abuse, and online content).

This change would remove the existing distinctions in the BSA between hosting services, live content services, and links services, and their resulting separate compliance pathways by treating content the same way. Instead, the scheme would focus on harmful content where it appears on the services most used by Australians.

Take-down notices would be generic, rather than retaining the existing distinctions between take-down (for hosting services), service-cessation (for live content services) or link deletion (for links services) notices. These definitions are unnecessarily duplicative and complex, lead to the same content being treated in different ways, and providing multiple layers of obligation on the same service.

Using the more contemporary definitions currently used in the image-based abuse scheme will reduce duplication in a new Online Safety Act, add clarity, and make it easier for online service providers to comply with the requirements.

Figure 6: Proposed focus of a future online content scheme



New Online Content Scheme
Part of a new Online Safety Act

Class 1 Content: Seriously harmful material such as child sexual abuse material, abhorrent violent material, incitement to violence, other seriously harmful material as determined by legislative instrument

Class 2 Content: Content that would be classified as RC, X18+ and MA15+ under the National Classification Code (e.g. ranging from pornography, high impact, realistically simulated sex and/or violence down to coarse language)

Take-down powers

Take-down notices, referral to law enforcement and international networks where sufficiently serious

Ancillary Service Scheme

Limits access when take-down notices are not effective

Blocking powers

For online safety crisis events

Industry code

Mandatory requirement that RC and X18+ content must not be hosted in Australia.

Require use of best available technology to prevent children's access to harmful content.

Developed in consultation with the community, approved by eSafety Commissioner.

Industry standard

Made by eSafety Commissioner if industry code is not made or is insufficient

Questions

19. Is the proposed application of the take-down powers in the revised online content scheme appropriate?
20. Are there other methods to manage access to harmful online content that should be considered in the new Online Safety Act?
21. Are there services that should be covered by the new online content scheme other than social media services, relevant electronic services and designated internet services?
22. Is the proposed take-down period of 24 hours for the online content scheme reasonable or should this require take-down in a shorter period of time?
23. Which elements of the existing co-regulatory requirements should be retained under the new Act?

Opt-in tools and services to restrict access to inappropriate content

Many parents are concerned about their ability to protect their children from online harms and feel powerless to identify the best ways of doing this. Providing people with the option to choose tools and services to filter content they do not wish to see would empower them and encourage a user-focused approach to online safety.

The Government has committed during the 2019 federal election to making available to parents and carers the option of a filtered internet service that, at a minimum, blocks access to websites identified by the eSafety Commissioner.

Current approach

Under the current co-regulatory arrangements, there are two Family Friendly accreditation processes that apply to online service providers.

1. Accredited family friendly providers – ISPs, content hosts and mobile carriers who are compliant with the relevant sections of the industry codes may be eligible to signify their compliance with the Family Friendly Program by placing the Ladybird Seal on their Safety Page and their products and services.[68]
2. Accredited family friendly filters – under the Content Code 3 of the Internet Industry Codes of Practice – Internet and Mobile Content, an ISP must make available one or more accredited filter.[69]

Need for change

The Family Friendly Filter Scheme is not widely known and does not capture the diverse range of Internet-connected devices now available in homes, nor the range of online services available.

As noted in the previous section, it is proposed that a revised industry code under a new Online Safety Act would include the requirement for online service providers to use the best available technology solutions to help Australian families to limit access to harmful content, whether it is hosted in Australia or offshore.

There are a wide range of tools and services available to assist users to safely manage their engagement online, as outlined in the table below. However, there is no externally validated way for users to compare these tools and services and determine whether they are effective.

[68] https://www.commsalliance.com.au/Activities/ispi/ffisp

[69] https://www.commsalliance.com.au/__data/assets/pdf_file/0003/44607/Internet-Industrys-Code-of-Practice-Internet-and-mobile-content-ContentCodes10_4.pdf

Table 2: Overview of some available tools and services

Service Provider	Tools
Mobile service providers	Telstra has a network-based solution available – *Smart Controls* – which allows account holders to set restrictions on calls, SMS, MMS and internet browsing for any mobile service included on the account. Optus and Vodafone have developed their own applications (apps) for mobile devices: the Optus *Mobile Security App*, available for Apple and Android devices, allows parents to set restrictions on apps and install safe browsing. The Vodafone *Guardian* App (for Android) allows parents to set restrictions on calls, SMS, other apps, internet browsing, and time of day controls.
Network-based controls	While most ISPs provide general information to keep account-holders safe online, Telstra offers Telstra Broadband Protect – a network-level filtering option that works with all internet connected devices. The product offers protection from spyware and viruses, but also blocks access to inappropriate websites such as those showing pornography or violence.
Device level controls	Parental controls are now available on most connected devices, including computers, tablets, smartphones, and gaming consoles. These controls may limit screen time or play time, or block access to specific sites or search term results such as pornography. Some tool settings can be tailored based on the child's age and skills. This grants parents and carers the ability to choose settings that suit their parenting needs.
Digital distribution platforms	Mobile app stores (e.g. Google Play Store, Apple Store) and video game distribution platforms (e.g. Steam, Nintendo eShop) offer control settings to restrict a child's ability to view catalogued items that are targeted at adults, and restrict a child's ability to make unauthorised purchases.
Web browsers	Users may choose to install browser add-on software, or change settings, to enable children to more safely explore the internet.

Service Provider	Tools
Child-friendly services	Some online service providers have developed child-friendly services such as: › Search engines, such as Kiddle, Kidtopia and KidsSearch.com › YouTube Kids › Social networks[70] – Targeted at 6–10 years old: Kudos, Playkids Talk, ChatFoss – Targeted at 11–13 years old: Kidzworld, Popjam – Targeted at 14+ years: Grom Social.

In addition to these tools and services, new technologies are emerging that could also play a role in restricting access to inappropriate online content. For example, the UK has examined – although recently shelved plans – to introduce an age-verification scheme to restrict access to online pornography to users over 18 years of age. Under the scheme, websites would have been required to implement technological solutions to allow people to prove they were over 18, with checks to be carried out by the UK's film regulator. A range of companies developed age verification systems using various methodologies including those that verified identity documents either in person or online, or using age estimation technology using facial images.

In September 2019, the House of Representatives Standing Committee on Social Policy and Legal Affairs commenced an inquiry into age verification tools for online wagering and online pornography.[71] The outcomes of this inquiry will be considered by the Government in implementing this proposal.

Proposal

Requirements to provide information about online safety tools and services

As outlined in the previous section, it is proposed that a new Online Safety Act would include industry code making powers to update the current industry code arrangements to require service providers to use the best available technology to prevent children's access to harmful content.

The codes would also include requirements for service providers to make available to consumers information relating to opt-in tools and services in order to educate users about the steps they can take to manage their own online safety.

[70] https://www.internetmatters.org/resources/social-media-networks-made-for-kids/

[71] https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification

Upgrades to these tools and services, and the impact of significant software upgrades, should also be communicated regularly to promote good online safety habits beyond point-of-purchase. This could be achieved by requiring service providers to link to a designated online safety page, or to the relevant resources of the eSafety Commissioner.

An accreditation scheme run by the eSafety Commissioner

A number of the online safety initiatives by industry have proven to be valuable to Australian users. However, the diversity of tools and services can be confusing to users trying to determine the best way of protecting children and other vulnerable people from inappropriate material. Factors to consider include:

- gaps in knowledge around the coverage of certain tools and services and where users will need additional tools or services to reinforce protections;
- sustainability in terms of financial cost and effort to maintain multiple tools and services on various devices and platforms; and
- lack of messaging or information around good online safety practices such as reminders to review parental controls.

As such, there is likely to be benefit for consumers in promoting a baseline of accessible and affordable opt-in tools and services. This could be achieved through an accreditation program to evaluate the mainstream tools and services available in the market. Information about accredited tools and services could be made available on the eSafety Commissioner and industry websites. The eSafety Commissioner would be expected to have an oversight role of this scheme. The accreditation would need to be tested periodically, ideally annually. There would be a need for ongoing funding for the eSafety Commissioner for this program to be effective.

An accreditation scheme run by an industry body

As an alternative, an industry body such as Communications Alliance could develop an accreditation scheme for use by members. Participation in the scheme would be at industry cost, rather than the Government's, but accreditation could be a useful tool for marketing to parents seeking to protect children from online harms.

Questions

24. To what extent would an expanded accreditation scheme for opt-in tools and services assist parents and carers in mitigating the risk of access by minors to potentially harmful content?
25. What categories of tools and services should be included in an accreditation program, aside from content filters?
26. What are the likely costs of developing and maintaining an accreditation scheme for opt-in tools and services to assist parents and carers in managing access to online content by minors?
27. When evaluating opt-in tools and services for accreditation, what criteria should be considered?

Blocking measures for terrorist and extreme violent material online

Current arrangements

The live-streamed terrorist attacks in Christchurch, New Zealand on 15 March 2019, and other subsequent events such as the shootings in Halle, Germany in October 2019, have demonstrated the risk of online services and platforms being used to promote violent and extremist actions. This can both compound the harm experienced by the victims of such actions, and contribute to the radicalisation of end-users.[72]

In the immediate aftermath of the Christchurch attacks, major ISPs in Australia voluntarily blocked access to sites known to contain footage of the attacks and the manifesto of the alleged perpetrator. ISPs blocked access to complete domains rather than individual URLs. This meant much inoffensive material could not be reached. This action, which prevented a great many people being exposed to online harm, also attracted criticism as there was no regulatory requirement to block the sites.

The Taskforce to Combat Terrorist and Extreme Violent Material Online established by the Government in the wake of the Christchurch attacks recommended that the Government pursue legislative change to establish a clear content blocking framework for terrorist and extreme violent material in online crisis events. An online crisis event was defined in the report of the Taskforce as 'an event that involves terrorist or extreme violent material being disseminated online in a manner likely to cause significant harm to the Australian community, and that warrants a rapid, coordinated and decisive response by industry and relevant government agencies'.

As an interim step, the eSafety Commissioner utilised a power under subsection 581(2A) of the *Telecommunications Act 1997* to issue a direction to Australian ISPs to continue to block eight domains still containing the Christchurch material.[73] This cemented the measures already in place to protect Australians from exposure to this material. However, there are limitations and shortcomings with the use of this power.

This power is not specifically directed or contained to blocking terrorist or extreme violent content, but is rather a broad power for the eSafety Commissioner to issue directions to a carrier or service provider in connection with the performance of the eSafety Commissioner's functions or the exercise of powers. The broad nature of this power has drawn criticism for being too open, and has led to concerns about its potential impacts on free expression.

[72] https://www.esafety.gov.au/esafety-information/-/media/cesc/sbd/safety_by_design_overview.pdf p. 5

[73] Telecommunications (Protecting Australian's from Terrorist or Violent Criminal Material) Direction (No 1) 2019

The current blocking arrangements do not provide civil immunity for ISPs when acting in accordance with a blocking direction. For these reasons, the report of the Taskforce recommended the establishment of a content blocking framework for terrorist and extreme violent material during an online crisis event.[74]

For website blocking to be an effective tool to support online safety outcomes in crisis events, the blocks need to be put in place quickly to counter the virality of this content. Currently it takes some time, normally days, before an instrument can be made and registered (and take effect). This limits its effectiveness in supporting online safety outcomes during a crisis event.

Proposals

The new Online Safety Act would establish a specific and targeted power for the eSafety Commissioner to direct ISPs to block certain domains containing terrorist or extreme violent material, for time limited periods, in the event of an online crisis event. The use of the power would be limited to dealing with online crisis events that involve terrorist or extreme violent material. This power would not be available to block websites on a routine or ongoing basis. It is proposed that the new power would also:

- provide ISPs with civil immunity from any action or other proceeding for damages as a result of implementing the requested blocks, (mirroring the arrangements under section 313 of the Telecommunications Act);
- require the eSafety Commissioner to notify owners of affected domains that their services had been blocked, and provide for appropriate appeal and review mechanisms; and
- stipulate that the arrangements for the blocks should be automated to the fullest extent possible and enable ISPs to implement and maintain any required blocks without the need for dedicated staff.

The eSafety Commissioner would be required to develop a protocol for the use of the new power. This protocol would set out the arrangements and processes for implementing blocks of websites hosting offending content, including:

- the means of determining which ISPs would be subject to blocking orders, the length of time that the ISPs will be required to implement the blocks, and the process for removing the blocks;
- the processes to be used to determine whether the terrorist or extreme violent material is sufficiently serious to warrant blocking action and to identify the domains that are hosting the material;
- guidance on the circumstances in which it is anticipated that this power may be used by the eSafety Commissioner; and
- the landing page for the blocked domains and the method of communicating the notice.

[74] Report of the Australian Taskforce to combat terrorist and extreme violent material online, 30 June 2019 https://www.pmc.gov.au/resource-centre/national-security/report-australian-taskforce-combat-terrorist-and-extreme-violent-material-online. Recommendations 5.2 and 5.3.

The proposal would also provide the eSafety Commissioner with the capacity to issue notices to ISPs. This would allow the eSafety Commissioner to act more quickly in responding to an online crisis event than the timeframes for making a legislative instrument permit.

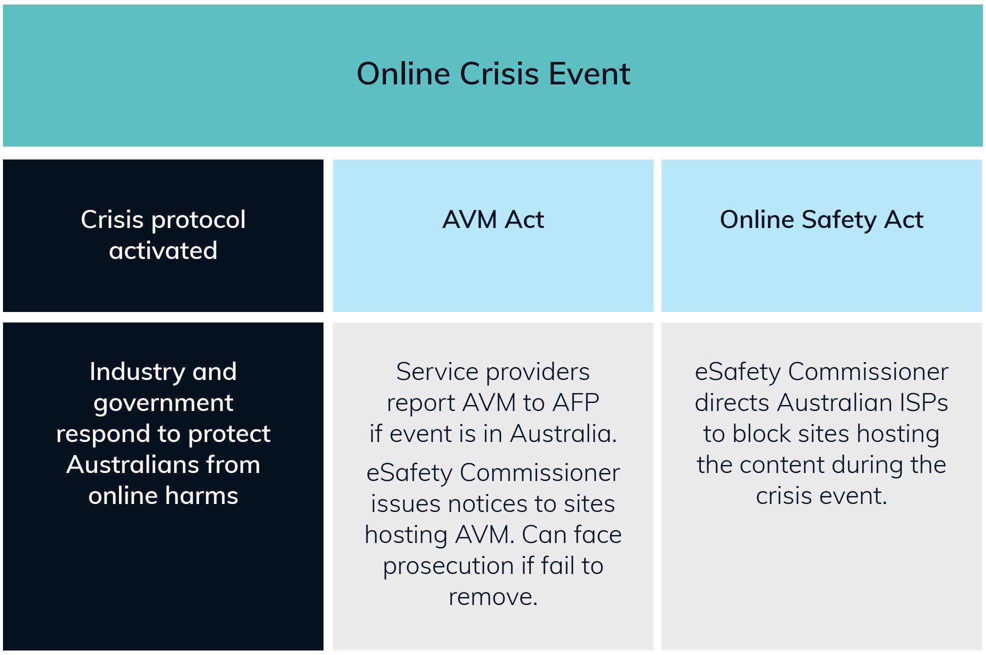
In the first instance, the eSafety Commissioner would issue voluntary notices. ISPs would not be required to respond to a voluntary notice, nor would there be any sanctions for non-compliance. ISPs would be provided with immunity from any civil liability for, or in relation to, an act done by an ISP in compliance with a notice.

However, if there was a need for further action, the voluntary notice scheme would be backed up with a power for the eSafety Commissioner to make mandatory notices that would require action by ISPs. These mandatory notices would be supported by compliance obligations and an enforcement mechanism through the Federal Court. As with the voluntary notices, immunity from civil liability would be provided to ISPs acting in compliance with a notice. The notices would be subject to appropriate appeals, transparency and oversight arrangements to ensure the proper and appropriate use of the power.

This approach – coupling a voluntary notice scheme (as a first point of call for the eSafety Commissioner) with a mandatory notice scheme (to be used only as required) – balances the need to need to act rapidly to address online safety concerns during online crisis events with broader principles of freedom of expression.

This blocking mechanism would complement the arrangements in the *Criminal Code Act 1995* for addressing AVM. The relationship between the two schemes is shown below.

Figure 7: Relationship between content blocking and AVM Act



Online Crisis Event		
Crisis protocol activated	AVM Act	Online Safety Act
Industry and government respond to protect Australians from online harms	Service providers report AVM to AFP if event is in Australia. eSafety Commissioner issues notices to sites hosting AVM. Can face prosecution if fail to remove.	eSafety Commissioner directs Australian ISPs to block sites hosting the content during the crisis event.

Questions

28. Is the proposed scope of content blocking for online crisis events appropriate?

29. Are there adequate appeals mechanisms available?

30. What other elements of a protocol may need to be considered?

Ancillary service provider notice scheme

Current arrangement

Harmful online content and conduct is technology-agnostic. In cases of serious cyberbullying and cyber abuse, image-based abuse, and creation and distribution of illegal and harmful online material, perpetrators can use multiple online services to cause harm.

The scope of services covered through a new Online Safety Act should reflect the online ecosystem, including the roles that differing services play in shaping a safe online environment. It should also be broad enough that the eSafety Commissioner is able to take multiple and appropriate courses of action to give Australians quick relief from harmful material.

Under the current arrangements, there is limited capacity for the eSafety Commissioner to seek assistance from service providers that are not directly responsible for the publication of harmful content or conduct, or who provide services that enable users to post harmful content or conduct.[75] The major gaps in service coverage relate to search aggregators and digital distribution platforms. While these services do not actively facilitate content creation, they are conduits for consumers to discover and access internet content and engage online.

- **Search aggregator services:** services that allows users to locate and access online content that is not hosted by the search aggregator or any related parties. For example, Google and Bing.
- **Digital distribution platforms:** services that host 3rd party services or products that end-users can download or otherwise use to access online content. For example, Google Play, Apple's App Store, Steam, GOG.

Proposals

A new ancillary service provider notice scheme would create a new service category of 'ancillary service provider', and enable the eSafety Commissioner to request (not require):

- search aggregator services to delist or de-rank websites that have been found by the eSafety Commissioner to be systemically and repeatedly facilitating the posting of cyberbullying or cyber abuse material, image-based abuse or hosting seriously harmful content; and

[75] Under the image-based abuse scheme, the eSafety Commissioner has the capacity to issue a removal notice to a hosting service provider.

- digital distribution platforms to cease offering apps or games found by the eSafety Commissioner be systemically and repeatedly facilitating the posting of cyberbullying or cyber abuse material, image-based abuse or hosting illegal or harmful content.

The Act would specify that these powers are intended to be used as 'reserve powers' in relation to ancillary service providers where more direct take-down powers used against the primary providers of harmful material have not been effective. This may well be the case for overseas hosted material or services, where the providers may have little regard for Australian laws.

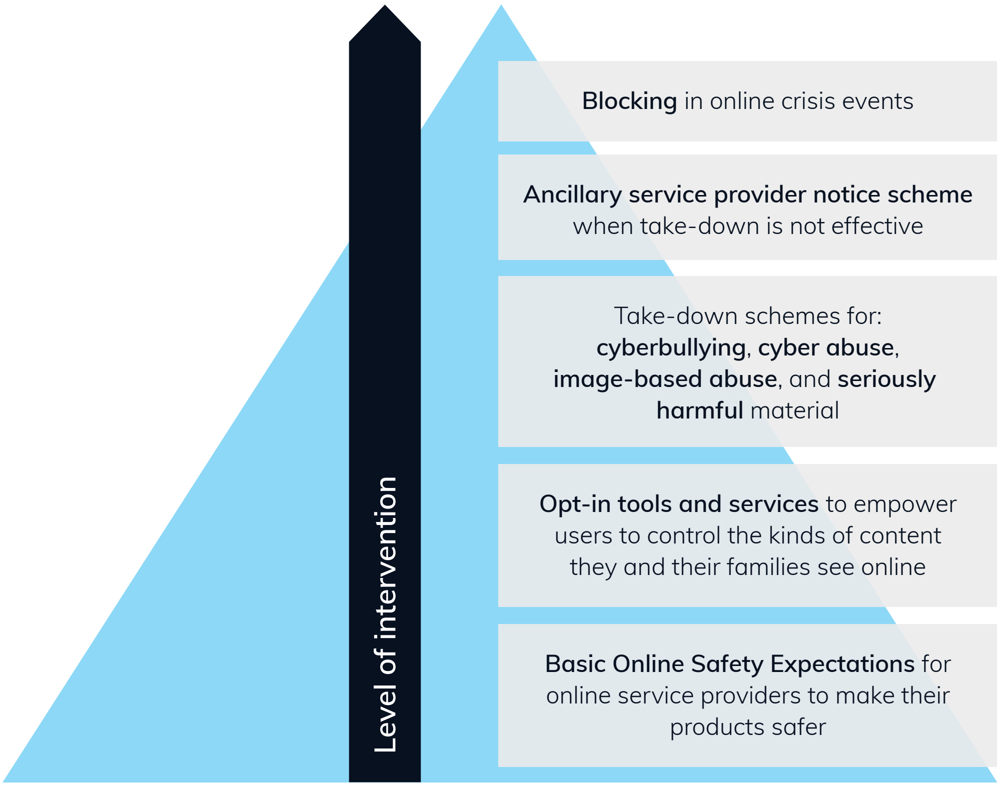
There would be no sanctions for non-compliance with an ancillary service provider notice, but the eSafety Commissioner would be empowered to publish reports on service providers who had failed to respond, or failed to respond adequately, to a notice. In order to capture emerging services under this scheme, the Act would include a provision that the Minister may determine, by legislative instrument, to exempt or include ancillary service providers.

These notices would provide the eSafety Commissioner with the scope to disrupt access to services that systemically provide, or provide a means to host, harmful online material. For the vast majority of cases this notice-making power would not need to be used; the eSafety Commissioner has demonstrated very high compliance rates when requesting the removal of cyberbullying material and image-based abuse at the source.

The scheme is intended to be proportionate and appropriate to the specific roles these services play in the online ecosystem. The function of the proposed scheme, with compliance with a notice being voluntary, acknowledges that measures that request search providers to delist and de-rank content need to be carefully and specifically constructed to avoid unreasonable impacts on free speech. The regulatory burden placed on ancillary services providers also needs to be appropriate and proportionate to the service they provide and their level of control over the content and conduct occurring on the underlying 3rd party-provided services.

Figure 8 illustrates how the various taken down and notification obligations operation in a scale from reporting to mandatory removal.

Figure 8: Visualisation of content removal and notification measures



Blocking in online crisis events

Ancillary service provider notice scheme when take-down is not effective

Take-down schemes for: **cyberbullying**, **cyber abuse**, **image-based abuse**, and **seriously harmful** material

Opt-in tools and services to empower users to control the kinds of content they and their families see online

Basic Online Safety Expectations for online service providers to make their products safer

Questions

31. Is there merit in the concept of an ancillary service provider notice scheme?
32. Are there any other types of services that should be included in the definition of ancillary service provider?
33. Should the definition of search engine provider be broadened to include search functions housed in other services, such as social media services, video hosting services or other services with internal search functionality?
34. Is the requirement that 3rd parties be systemically and repeatedly facilitating the posting of cyberbullying or cyber abuse material, image-based abuse or hosting illegal or harmful content appropriate before the eSafety Commissioner can issue a notice to an ancillary service provider? Should a different threshold be contemplated?
35. Is there merit to making compliance with the ancillary service provider notices mandatory?

Role of eSafety Commissioner

Current arrangements

The eSafety Commissioner was originally established in 2015 as the Children's eSafety Commissioner with a regulatory role limited to administering the new cyberbullying scheme and taking over the ACMA's function of administering the online content scheme. The eSafety Commissioner was established as an independent statutory office holder, supported by and operating within the ACMA.

Since that time, the Government expanded the responsibilities of the eSafety Commissioner to include online safety for all Australians and the administration of the image-based abuse scheme and notice-making powers in relation to abhorrent violent material. The proposals canvassed in this paper would significantly expand the eSafety Commissioner's remit and responsibilities. A summary of these proposed new responsibilities and powers for the eSafety Commissioner are included at **Attachment B**.

The functions and powers of the eSafety Commissioner are also listed in different pieces of legislation including the BSA, EOSA and Criminal Code Act. As the 2018 Review noted, the consequence of this fragmentation is that the full range of functions are not clear.[76] Nor are the diverse range of functions listed in Section 15 of the EOSA prioritised.

Proposals

Functions of the eSafety Commissioner

A new Act would consolidate and streamline the functions as far as possible to reflect the proposed objects of the new Act to:

- protect Australians online;
- promote online safety; and
- prevent online harms.

Some specific functions would still need to be maintained, including the power to make grants, along with specific regulatory functions under the cyberbullying, online content and image-based abuse schemes. New functions would also be considered, including those in relation to international engagement.

The eSafety Commissioner's current explicit functions are provided in Table 3, and comment is sought on the merits of consolidating and organising these functions to align with the proposed objects of the new Act.

[76] Ibid 27-28

Table 3: Current functions of the eSafety Commissioner

Current functions in EOSA
15 (1) (a) such functions as are conferred on the Commissioner by: (i) this Act; or (ii) Schedules 5 and 7 to the *Broadcasting Services Act 1992*; or (iii) any other law of the Commonwealth
(b) to promote online safety for Australians
(c) to support and encourage the implementation of measures to improve online safety for Australians
(d) to coordinate activities of Commonwealth Departments, authorities and agencies relating to online safety for children
(e) to collect, analyse, interpret and disseminate information relating to online safety for Australians
(f) to support, encourage, conduct, accredit and evaluate educational, promotional and community awareness programs that are relevant to online safety for Australians
(g) to make, on behalf of the Commonwealth, grants of financial assistance in relation to online safety for Australians
(h) to support, encourage, conduct and evaluate research about online safety for Australians
(i) to publish (whether on the internet or otherwise) reports and papers relating to online safety for Australians
(j) to give the Minister reports about online safety for Australians
(k) to advise the Minister about online safety for Australians
(l) to consult and cooperate with other persons, organisations and governments on online safety for Australians
(m) to advise and assist persons in relation to their obligations under this Act
(n) to monitor compliance with this Act
(o) to promote compliance with this Act

Current functions in EOSA
(p) to formulate, in writing, guidelines or statements that: (i) recommend best practices for persons and bodies involved in online safety for Australians; and (ii) are directed towards facilitating the timely and appropriate resolution of incidents involving cyber-bullying material targeted at an Australian child;
(q) to promote guidelines and statements formulated under paragraph (p)
(r) such other functions (if any) as are specified in the legislative rules
(s) to do anything incidental to or conducive to the performance of any of the above functions
Current functions conferred by legislative rule[77]
to promote online safety for Australians by protecting Australians from access or exposure to material that promotes, incites, or instructs in, terrorist acts or violent crimes.

Governance arrangements

The 2018 Review found that the governance arrangements for the Office of the eSafety Commissioner needed to be reformed given the expanding role and growth of the organisation since its inception. In particular, the 2018 Review recommended that the eSafety Commissioner become part of the Department of Communications and the Arts as part of a transition to a fully independent organisation.[78] The Government is considering this and other models for governance reform, including:

1. Establish the eSafety Commissioner **as a separate, standalone Commonwealth entity**.
   - Under this option, the eSafety Commissioner would continue to be an independent statutory office holder, but with full control over and responsibility for all aspects of the organisation similar to regulators such as the ACMA and the ACCC.
2. Merge the eSafety Commissioner with **a different department or agency permanently**.
   - Under this option, the eSafety Commissioner would continue to be an independent statutory office holder, supported by staff from a relevant department or agency.

[77] *Enhancing Online Safety (Protecting Australians from Terrorist or Violent Criminal Material) Legislative Rule 2019*

[78] L. Broggs op. cit p.37

3. **No change to governance arrangements**, but clearer powers for the organisation.
   - Under this option, the eSafety Commissioner would continue to be an independent statutory office holder with staff and corporate support provided from the ACMA. The new Online Safety Act would expand the functions, and clarify the powers, of the eSafety Commissioner.

The Government is evaluating these options and the likely costs of any ancillary impacts on government

It is likely that an expanded regulatory power for the eSafety Commissioner would have impacts on other parts of government. For example, the expansion in the eSafety Commissioner's range powers to issue notices might lead to an increased number of actions to review of these notices in Federal tribunals such as the Administrative Appeals Tribunal and the Federal Court. While the potential number of actions is difficult to gauge, there may be resource implications for these agencies as well as for businesses who choose to challenge a notice. These issues will be considered by the Government in the development of the new Online Safety Act.

Questions

36. Are the eSafety Commissioner's functions still fit for purpose? Is anything missing?
37. To what extent should the existing functions of the eSafety Commissioner be streamlined? Are there particular functions that need to be maintained, or new functions that should be specified?
38. To what extent should the functions of the eSafety Commissioner be prioritised?
39. What are the likely impacts, including resource implications, on other agencies and businesses of a new Online Safety Act?

Attachments

Attachment A – Powers, civil penalties and offences under the current online safety schemes

The *Broadcasting Services Act 1992* and the *Enhancing Online Safety Act 2015* provide for a graduated range of enforcement tools for the eSafety Commissioner, including informal and formal warnings, remedial directions, enforceable undertakings and injunctions. These enforcement tools are canvassed in the relevant sections of this discussion paper. The following table outlines some key powers and civil penalties available to the eSafety Commissioner under the cyberbullying, image-based abuse and online content schemes.

Scheme	Key powers	Maximum penalties
Cyber-bullying scheme	The eSafety Commissioner may **request Tier 1**[79] social media services to remove cyberbullying material. The eSafety Commissioner can issue a **social media service notice**, requiring the removal of cyberbullying material within 48 hours of being given the notice, to **Tier 2** social media services. The eSafety Commissioner can issue an **end-user notice** to an individual user to require the removal of cyberbullying material from a social media service or relevant electronic service.	There is no civil penalty, but the service can have its **Tier 1** status revoked, making it a **Tier 2** service Failure to comply with a social media service notice: a civil penalty of up to **$21,000 for individuals and up to $105,000** for bodies corporate) End-users face no civil penalties, but can have an injunction taken out against them to enforce compliance.

[79] Tier 1 services are those social media services which have applied to the eSafety Commissioner for this status and satisfied the eSafety Commissioner that they comply with basic online safety requirements set out in s 21 of the EOSA. Tier 2 services are declared as such by the Minister following a recommendation from the eSafety Commissioner, for example, because they have failed to apply for tier 1 status and/or do not comply with the basic online safety requirements.

Scheme	Key powers	Maximum penalties
Image-based abuse scheme	The eSafety Commissioner can issue **removal notices** to providers and end-users of social media services, relevant electronic services, and designated internet services; and to hosting service providers. The eSafety Commissioner can give a person a **remedial direction**.	<u>**Service providers**</u> Failure to comply with a removal notice: civil penalty up to **$105,000** for individuals and up to **$525,000** for bodies corporate). Failure to comply with a remedial direction: civil penalty up to **$105,000** for individuals and up to **$525,000** for bodies corporate). <u>**End-users**</u> Posting, or threatening to post: civil penalty of up to **$105,000** for individuals and up to **$525,000** for bodies corporate). Failure to comply with a removal notice: civil penalty of up to **$105,000** for individuals and up to **$525,000** for bodies corporate).[80]

[80] Persons can also face prosecution under the *Criminal Code 1995*, for aggravated offences with prison terms of up to five years for using a carriage service in a way that reasonable persons would regard as being menacing, harassing or offensive, where the commission of that offence involves the transmission, making available, publication, distribution, advertisement or promotion of private sexual material. Persons can also face imprisonment for up to seven years in cases where prior to the commission of that offence, three or more civil penalty orders were made against the person for posting or making a threat to post an intimate image of another person in contravention of subsection 44B(1) of the *Enhancing Online Safety Act 2015*.

Scheme	Key powers	Maximum penalties
Online content scheme – content hosted in Australia	The eSafety Commissioner can issue notices to certain service providers in relation to prohibited or potentially prohibited content in Australia, to require them **to take down content, cease services or to delete links**. The eSafety Commissioner can give a person a **remedial direction**.	A failure to comply with a notice or a remedial direction is an offence of up to **$21,000** for individuals and up to **$105,000** for bodies corporate, and is a civil penalty provision. It is a separate offence/ contravention for each day during which the contravention continues.
Online content scheme – content hosted outside Australia[81]	If there is a relevant industry code, the eSafety Commissioner notifies prohibited or potential prohibited content to Australian Internet Service Providers (ISPs) under the 'designated notification scheme' in the code.[82] The eSafety Commissioner can direct the ISP to comply with a registered industry code. If there is no code, the eSafety Commissioner can issue ISPs with a 'standard access-prevention notice' and if needed, a 'special-access prevention notice'.	Failure to comply with a direction to comply is an offence of up to **$10,500** for individuals and **$52,500** for bodies corporate.). It is a separate offence for each day of contravention. Failure to take all reasonable steps to prevent an end-user accessing prohibited or potentially prohibited content is a criminal offence, of up to $10,500 for individuals and up to **$52,500** for bodies corporate).

[81] There are no powers in the BSA to issue takes down notices to overseas sites.

[82] An industry code is currently in place.

Attachment B – Table showing proposed new powers under the Online Safety Act

Powers	Significance of change
Basic online safety expectations Minister to have a power to specify, **via a legislative instrument, a set of basic online safety expectations (BOSE)**. eSafety Commissioner to have a power to determine, by legislative instrument, that particular entities report on their actions in upholding the BOSE, through public reporting and/or reporting on specific items to the eSafety Commissioner.	Significant expansion of the basic online safety requirements for social media services under the existing cyberbullying scheme
Cyberbullying eSafety Commissioner to have a power to require the removal of material within 24 hours (rather than 48 hours).	Minor change for consistency with other schemes
Cyber abuse of adults eSafety Commissioner to have a new power to address sufficiently serious cyber abuse of adults with: › power to require removal of material within 24 hours and › a civil penalty regime for perpetrators of cyber abuse.	New scheme
Non-consensual sharing of intimate images eSafety Commissioner to have a power to require the removal of material within 24 hours (rather than 48 hours).	Minor change for consistency with other schemes

Powers	Significance of change
Online content scheme	
eSafety Commissioner to be empowered to determine if particular content is seriously harmful.	Change from reliance on the Classification Board
eSafety Commissioner to have a power to order the removal of seriously harmful material within 24 hours.	Minor change for consistency with other schemes
eSafety Commissioner to be able to order the take-down of seriously harmful material.	Significant expansion of role.
Minister to have the power on the basis of the advice of the eSafety Commissioner to have the power to determine by legislative instrument that new types of content are seriously harmful.	New power
Ancillary service provider notice scheme	
eSafety Commissioner to have new reserve powers to request search engines de-rank websites, and digital distribution platforms to cease offering games or apps, that facilitate access to harmful material.	New power
eSafety Commissioner empowered to publish reports on service providers who had failed to respond, or failed to respond adequately, to a notice.	New power
Minister to have the power to determine by legislative instrument to exempt or include ancillary service providers in the scheme	New power
Accreditation scheme for opt-in tools	
eSafety Commissioner to oversight an accreditation program to evaluate the mainstream tools and services available in the market.	New responsibility for the eSafety Commissioner which would be resource intensive.

Powers	Significance of change
Blocking measures for terrorist and extreme violent material online The eSafety Commissioner to direct ISPs to block certain domains containing terrorist or extreme violent material, for time limited periods, in the event of an online crisis event. The use of the power would be limited to dealing with online crisis events that involve terrorist or extreme violent material.	New power to implement a recommendation of the Taskforce to Combat Terrorist and Extreme Violent Material.