Online Safety Act 2021

Development of industry codes
under the Online Safety Act
Position Paper

September 2021

Australian Government | eSafetyCommissioner

Acknowledgement

eSafety acknowledges the Traditional Custodians of Country throughout Australia and their continuing connection to land, waters and community. We pay our respects to Aboriginal and Torres Strait Islander cultures, and to Elders past, present and emerging.

Given the global nature of the internet, eSafety also acknowledges the inherent and continuing rights of indigenous people across the globe.

Contents

Executive summary

The new Online Safety Act 2021 (Cth) (the Act) will commence on 23 January 2022. The Act provides for industry bodies or associations to develop, and eSafety to register, new industry codes to regulate harmful online content. This material, referred to as 'class 1' and 'class 2' material, ranges from material of the highest and most serious harm, such as videos of the sexual abuse of children or terrorism, through to material which is inappropriate for children, such as online pornography.

This paper includes a series of positions that outline the eSafety Commissioner's (eSafety) expectations for the development of codes, as well as its preferred outcomes-based model for the codes. In order to register a code, eSafety expects the codes to align, as much as possible, with the positions and model set out in this paper.

eSafety's expectations to guide codes development

The 11 positions outlined in this paper cover threshold issues about the substance, design, development, registration and administration of the codes, and are intended to assist industry bodies and associations to prepare codes. These positions are:

Substance	1. The codes will address the issues of access, exposure and distribution that are related to class 1 and class 2 material 2. The application of the codes will not be limited to services provided from Australia
Design	3. Industry associations will develop a set of common drafting principles to inform codes development 4. The codes will adopt an outcomes- and risk-based regulatory approach supported by clear compliance measures which apply to industry participants whose services or devices present the greatest risk in respect of class 1 and class 2 material
Development and registration	5. Industry associations will prepare all codes for registration by July 2022 or adopt a phased approach to codes development. Under the phased approach, codes dealing with the most harmful content must be lodged for registration by July 2022, and codes dealing with content which is inappropriate for children must be lodged for registration by December 2022 6. Industry associations will limit the number of codes developed. 7. Industry associations will engage widely with participants within their industry section(s) to ensure they adequately represent each industry section covered by a code 8. Industry associations will conduct meaningful industry and public consultation 9. Industry associations will engage with eSafety throughout the codes development process
Administration	10. Industry participants will handle reports and complaints about class 1 and class 2 material and codes compliance in the first instance.

eSafety will act as a 'safety net' if resolution of a complaint is not satisfactory

11. The codes will include a review mechanism

eSafety has developed a preferred codes model to assist industry participants in composing their codes

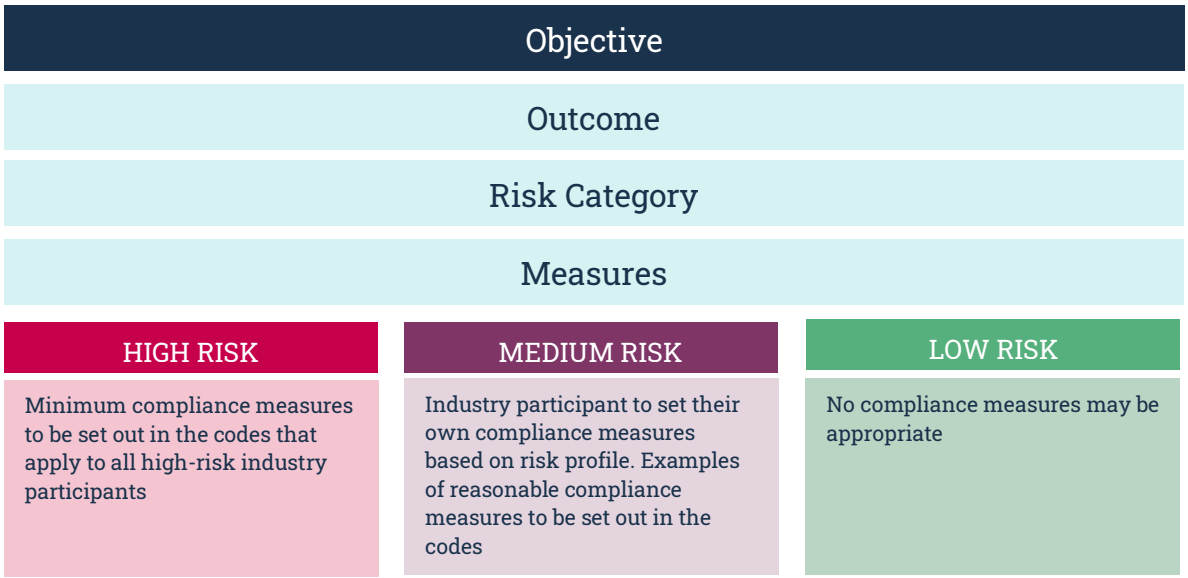
eSafety considers that industry associations should adopt an outcomes- and risk-based approach when developing the codes. This approach has been informed by a review of local and international regulatory approaches and preliminary discussions with industry.

An outcomes-based approach provides industry participants with a common set of objectives and outcomes, while granting the flexibility to implement measures to meet those objectives and outcomes that are most suited to their business models and technologies. These measures should be reasonable and proportionate, based on an assessment of the risk an industry participant's services or devices present in respect of class 1 and class 2 material.

eSafety has developed a codes model which sets out its preferred objectives and outcomes for the codes.

However, to ensure high-risk industry participants can be held to account, eSafety considers that the outcomes-based model should be supported by clearly defined minimum compliance measures for each outcome. These measures should apply to industry participants whose services and devices are assessed under the codes as presenting the greatest risk in respect of class 1 and class 2 material. eSafety expects industry to build minimum compliance measures into the outcomes-based model.

Suggested approach to codes



Preferred outcomes-based codes model

<table>
<tr><th colspan="4">PURPOSE: To ensure that participants of the online industry provide appropriate community safeguards for Australians in relation to class 1 and class 2 material</th></tr>
<tr><td colspan="4">OBJECTIVE 1: Industry participants will take proactive steps to create and maintain a safe online environment</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Material prevention or restriction</td><td>Industry participants proactively detect and prevent:*
access or exposure to,
distribution of, and
online storage of,
Class 1 - 1A[1] material</td><td>Industry participants proactively prevent or limit:
access or exposure to, and
distribution of,
Class 1 - 1B[2] material **</td><td>Industry participants proactively:
prevent access or exposure to, and distribution of, or
prevent children from accessing or being exposed to,
Class 1 - 1C[3] and Class 2[4] material ***</td></tr>
<tr><td>Hosting</td><td colspan="3">Industry participants do not host class 1 and class 2 – 2A[5] material in Australia. Industry participants who host Class 2 – 2B[6] material in Australia prevent children from accessing, or being exposed to, that material</td></tr>
<tr><td>Industry cooperation</td><td colspan="3">Industry participants consult, cooperate and collaborate with other industry participants in respect of the removal, disruption and/or restriction of class 1 and class 2 material.</td></tr>
<tr><td>Cooperation with Commissioner</td><td colspan="3">Industry participants communicate and cooperate with the eSafety Commissioner in respect of matters relating to class 1 and class 2 material, including complaints</td></tr>
<tr><td colspan="4">OBJECTIVE 2: Industry participants will empower people to manage access and exposure to class 1 and class 2 material</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Tools and information</td><td colspan="3">Industry participants provide tools and/or information to limit access and exposure to class 1 and class 2 material</td></tr>
<tr><td>Reporting of material</td><td colspan="3">Industry participants provide robust and effective reporting and complaints mechanisms for class 1 and class 2 material</td></tr>
<tr><td>Report handling</td><td colspan="3">Industry participants effectively respond to reports and complaints about class 1 and class 2 material</td></tr>
<tr><td colspan="4">OBJECTIVE 3: Industry participants will strengthen transparency of, and accountability for, class 1 and class 2 material</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Public policies</td><td colspan="3">Industry participants provide clear and accessible information about class 1 and class 2 material</td></tr>
</table>

[1] This includes child exploitation material, pro-terror content and extreme crime and violence.
[2] This includes crime and violence and drug-related content.
[3] This includes online pornography (RC).
[4] This includes online pornography (X18+), online pornography (R18+) and other high impact content.
[5] This includes online pornography (X18+).
[6] This includes online pornography (R18+) and other high impact content.

Public reporting	Industry participants publish periodic reports about class 1 and class 2 material and codes compliance

*Industry participants should take reasonable steps to proactively prevent access or exposure to, and distribution and online storage of, this material. However, failure to prevent access or exposure to, and distribution and online storage of, this material, does not necessarily indicate that there has been a codes breach. Where this material is accessible, 'prevention' includes quick removal of this material.

** At a minimum, industry participants must proactively limit access or exposure to, and distribution of, class 1 – 1B material.

***At a minimum, industry participants must proactively prevent children from accessing, or being exposed to, class 1 - 1C and class 2 material.

This paper addresses issues associated with the development of codes at a high level and provides industry with a foundation from which to commence drafting the codes. eSafety anticipates working closely and collaboratively with industry throughout the drafting process, and will consider all ideas, feedback and alternative approaches proposed by industry. However, industry will need to ensure that any deviations from the positions in this paper still achieve their policy intent and result in robust codes which offer meaningful protections for Australians in respect of class 1 and class 2 material.

1.0 Context and purpose

Overview

eSafety is Australia's independent online safety regulator. eSafety's mandate is to minimise online harms so Australians have safer, more positive online experiences. One of the ways in which eSafety achieves this mandate is by exercising its legislative powers to help prevent exposure to harmful online content and behaviour, and to investigate and remediate when complaints are made about such content or behaviour.

Harmful online content ranges from material of the highest and most serious harm, such as videos of the sexual abuse of children or terrorism, through to material which is inappropriate for children, such as online pornography.

Referred to as 'class 1' and 'class 2', this material is subject to the Online Content Scheme, which is being updated as part of the new Online Safety Act. The Act passed Parliament on 23 June 2021 and will become operational on 23 January 2022. The new Online Content Scheme requires that new industry codes or standards are developed to regulate these types of harmful online content, as the current codes are no longer fit for purpose – they are outdated, limited in scope, prescriptive and not capable of being sufficiently implemented or enforced.[7]

The current codes were developed under the Broadcasting Services Act 1992 (Cth) (BSA). In the almost 20 years since the first codes were registered, digital connectivity has changed dramatically. Smartphones have emerged and taken over as one of the most common forms of communication, the use of social media services has become prolific and online engagement is now ubiquitous. The rapid pace of technological change has brought us new platforms, services and functionality, including messaging apps, interactive games and live streaming. Meanwhile the volume and range of content available online has grown exponentially. These changes have brought great benefits, but also significant risks.

The time is right for a new approach.

Ideally, the online industry would play a critical co-regulatory role in Australia. Under this model, industry's peak bodies would draft reasonable and effective

[7] Lynelle Briggs AO, 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 1922 (Online Content Scheme), October 2018, https://www.infrastructure.gov.au/sites/default/files/briggs-report-stat-review-enhancing-online-safety-act2015.pdf.

codes that contain adequate mechanisms for preventing or limiting harmful online content.

eSafety believes that industry plays an important part in the online safety ecosystem and has the technical expertise and understanding to develop robust codes. However, if appropriate codes cannot be established, the eSafety Commissioner has the power under the Act to declare standards.

Purpose of this position paper

This position paper is intended to inform and guide the development of industry codes, while also facilitating ongoing engagement between eSafety and the online industry. eSafety hopes to foster a close and collaborative relationship that will result in codes the industry can comply with, eSafety can enforce and the wider community is willing to support.

The paper outlines the nature and impact of harmful online content, current approaches to regulating it and eSafety's expectations for regulation under the new Online Content Scheme.

It sets out 11 positions regarding the substance, design, development, registration and administration of industry codes. It proposes that industry associations adopt an outcomes- and risk-based approach when developing codes, supported by clear compliance measures which apply to industry participants whose services or devices present the greatest risk in respect of class 1 and class 2 material. eSafety has developed its preferred objectives and outcomes, which are set out in **5.0 Preferred codes model.**

When registering codes, eSafety will consider the extent to which they align with the 11 positions put forward in this paper, as well as the preferred codes model.

eSafety's operations

Established in 2015, eSafety is the first government agency in the world dedicated specifically to online safety. eSafety leads and coordinates online safety efforts across Australian Government departments, authorities and agencies. It also provides guidance for industry and the broader community so they can understand, prevent and respond to harms. As the internet has no borders and online safety is a global issue, eSafety also engages with key online safety stakeholders internationally to amplify its impact.

eSafety takes a whole of community approach to online safety, drawing on a range of social, cultural, technological and regulatory initiatives and interventions. Its complaints-handling experience, extensive research and rigorous evaluation provide a strong evidence base for our programs, resources and consumer advice. eSafety's work reflects its core pillars of protection, prevention and proactive and systemic change.

eSafety recognises that online risks are higher for some sections of the Australian community and that harms can have disproportionate impacts, especially for disadvantaged or marginalised people who have multiple, intersecting risk factors. This includes, but is not limited to, Aboriginal and Torres Strait Islander people, people from culturally and linguistically diverse communities, people with disability and people who identify as LGBTQIA+, as well as, depending on the circumstances, women, older people and children and young people. Children and young people are particularly at greater risk from the social, emotional, psychological and even physical impacts that can result from exposure to harmful content and behaviour online.

Apart from the Online Content Scheme, eSafety leverages its legal powers and online industry relationships to have other types of harmful content removed from the internet. This can happen in the case of serious cyberbullying of children that involves sharing abusive content and, for victims of any age, when intimate images or videos are shared without the consent of the person shown. Under the Online Safety Act, eSafety will also be able to intervene in cases of serious adult cyber abuse.

Codes under the Online Safety Act

Part 9, Division 7 of the Act allows for the establishment of new industry codes or standards. Codes are to be developed by industry bodies or associations, while eSafety is responsible for drafting and registering industry standards.

The codes or standards will apply to the participants of eight key sections of the online industry that provide a wide range of services. These include providers of social media, email, messaging, gaming, dating, search engine and app distribution services, as well as internet and hosting service providers, manufacturers and suppliers of equipment used to access online services and those that install and maintain the equipment.[8]

[8] See section 135 of the Online Safety Act.

eSafety will be able to receive complaints and investigate potential breaches of the codes or standards, and they will be enforceable by civil penalties, enforceable undertakings and injunctions to ensure compliance. However, industry participants should take responsibility for codes complaints in the first instance.

The Act provides an extensive list of examples of matters that may be dealt with by industry codes and standards. These matters broadly fit into three categories which align with eSafety's preferred objectives for the codes:

- Measures to create and maintain a safe online environment
- Measures to empower persons to manage access to class 1 and class 2 material
- Measures focused on transparency and accountability.

As Australia's expert and leader in online safety, eSafety is uniquely placed to oversee the development and implementation of industry codes in a way that best meets the safety needs of Australians.

Engagement and analysis

This position paper draws on eSafety's engagement with industry and is informed by a review of local and international regulatory approaches.

In developing the policy positions outlined in this paper, eSafety has taken a number of important steps, including:

- Engaging closely with industry bodies and associations, especially the Australian Mobile Telecommunications Association, BSA | The Software Alliance, Communications Alliance, Digital Industry Group Inc. (DIGI), the Internet Association of Australia and the Interactive Games and Entertainment Association (IGEA), as well as their members. eSafety would like to thank these associations and their members for the feedback, contributions and insights they have provided on the code process to date. eSafety also thanks the Consumer Electronic Suppliers Association and the National Retail Association for their recent engagement.
- Consulting with national regulators with interconnected regulatory schemes, including the Australian Communications and Media Authority (ACMA), the Office of the Australian Information Commissioner (OAIC), the Australian Competition and Consumer Commission (ACCC) and the Department of Home Affairs (Home Affairs). eSafety has also worked

closely with its portfolio department, the Department of Infrastructure, Transport, Regional Development and Communications.

- Analysing international approaches to online content and codes development.
- Reviewing the history of the Online Content Scheme and the experience of industry codes under the BSA. A timeline of online content regulation is set out at Appendix A.
- Carefully considering how industry codes can play a key role in Australia's multifaceted approach to online safety.

Terminology

Throughout this position paper, eSafety has used a number of terms that are explained in the following list:

- 'CSEM' refers to child sexual exploitation material. Based on the ECPAT Terminology Guidelines (also known as the Luxembourg Guidelines)[9], the term 'child sexual exploitation material' is a broad category of content that encompasses material that sexualises and is exploitative to the child, but that does not necessarily show the child's sexual abuse. Child sexual abuse material, which shows a sexual assault against a child, is a narrower category and can be considered a sub-set of CSEM. Class 1 material, which is defined under the Act by reference to the National Classification Scheme, includes material that is both sexually exploitative and that depicts or describes child sexual abuse.[10] For the sake of simplicity, this paper uses the term CSEM.
- 'Codes' refers to a single code or set of codes to be developed under the Act.
- 'Devices' refers to equipment that is for use by end-users in Australia of a social media service, relevant electronic service, designated internet service or internet carriage service, in connection with that service. While the Act uses the term 'equipment' when referring to persons who manufacture, supply, maintain or install this equipment, 'devices' is a more commonly understood term.

[9] Interagency Working Group on Sexual Exploitation of Children, Luxemburg Guidelines, January 2016, http://luxembourgguidelines.org/.

[10] More information about what constitutes CSEM can be found in the Guidelines for the Classification of Films. See also 2.0 Harmful online content for further discussion.

- ‘eSafety Commissioner’ refers to the statutory office of the eSafety Commissioner and is used when referencing the Commissioner’s functions and powers under the Act. ‘eSafety’ refers to the agency that undertakes the work of the eSafety Commissioner.
- ‘Industry’ refers to providers of services and manufacturers, suppliers, maintainers or installers of devices, as well as industry associations.
- ‘Industry association’ refers to an industry body or association who could draft a code.
- ‘Services’ refers to social media services, relevant electronic services, designated internet services, internet search engine services and app distribution services, so far as those services are provided to end-users in Australia, hosting services so far as those services host material in Australia and internet carriage services, so far as those services are provided to customers in Australia.
- ‘Victims’ refers to a victim or survivor, recognising that the choice of term is personal and important to the victim or survivor.

There is also a glossary at **Appendix B**, which provides the definitions for a number of the legislative terms that are referred to in this position paper.

2.0 Harmful online content

Types of online harms

Over recent decades, the rise of global digital platforms and services has resulted in shifts in the production, distribution and consumption of online content.

Online services and digital based technologies have provided vast benefits and unforeseen opportunities. But the same digital technologies that enable users to connect in important new ways can also be weaponised to perpetrate abuse.

A wide range of harms with multiple impacts on users can result from both online content and online behaviour.

Content harms can result from:

- the production of content – for example, where a perpetrator makes contact with a victim in an attempt to groom, coerce or force them into the production of content, or where coerced sexual activity or abuse is recorded
- the distribution of content – for example, where abusive material is posted, reshared or live-streamed online, which can compound the trauma experienced by victims harmed in the production of content
- the consumption of content – for example, where a person's behaviour, emotions, mental health, attitudes or perceptions are negatively impacted as a result of access or exposure to harmful content.

Harms can also result from online behaviours. Online environments can facilitate and amplify positive and negative interactions with others, both online and offline. This means the design, implementation and moderation of online environments plays an important role in the exposure of users to behavioural risks and harms that may affect their experiences.

Impacts of online harms

Harmful online content and behaviour can be seriously damaging, especially for those most at-risk, such as children and young people.

Online harms can violate a person's rights and their experiences and perceptions of their rights.

These include their right to:

- personal safety
- health and wellbeing
- dignity
- privacy
- participation and expression
- financial security
- truth and connection to reality.

The social, emotional, psychological and even physical impacts of online harms can be immediate, experienced over a period of time and/or enduring. They can also be experienced both online and offline.

It is important that the voices and experiences of victims are recognised and listened to when developing online safety solutions. The following quotes gathered by the Canadian Centre for Child Protection from survivors of child sexual exploitation underscore the deep and prolonged harm of CSEM.[11]

'The abuse stops and at some point also the fear for abuse; the fear for the material never ends.'

'The experiences are over. I can get a certain measure of control over those experiences. With regard to the imagery, I'm powerless. I can't get any control. The images are out there.'

'The images are indestructible and reach a huge lot of people and it is unstoppable. That's what makes it the worst thing for me. The idea that a complete and utter stranger has seen you and that I'm somebody's gratification right up to this very day.'

'Because the imagery continues to exist and you have no control over it. You never know who will see it. And if you get approached on the street by a

[11] Canadian Centre for Child Protection, Survivors Survey Full Report 2017, September 2017, https://protectchildren.ca/pdfs/C3P_SurvivorsSurveyFullReport2017.pdf.

total stranger who says "Don't I know you from somewhere?" or "You look familiar to me", you quickly link that to the imagery.'

Risk factors for individuals

The likelihood of a person experiencing online harm, and the level of harm they suffer, depends on many different factors.[12]

Personal factors that may place an individual at greater risk of experiencing or being seriously impacted by harmful online content or behaviour include low age and maturity, low digital literacy, lack of digital access, low self-esteem, cognitive development issues, anti-social behaviour, mental or physical illness and previous experiences of online harm.

Social factors that may increase a person's risk online relate to structural and systemic forms of inequality, discrimination and oppression. This includes sexism, racism, ableism, ageism, homophobia, biphobia and transphobia.

The factors can also be intersectional. This means that the layering of personal or social factors can increase a person's risk and impact on their online experiences at the same time, while different factors can impact on their risks and experiences at different times.

There is a strong link between the inequality, discrimination and disrespect that underpins harms experienced online and harms experienced offline.

This is why eSafety places a strong emphasis on capacity building for those most at-risk of online harm, such as children and those who support them, such as parents and carers.

Equipping users and support networks with information and advice so they can understand online risks, protect themselves with preventative behaviours, deal with safety issues as they arise and minimise the impacts of harm is a key part of eSafety's role. Accordingly, it will be essential for the new codes to include the provision of information and advice.

According to eSafety's research, the top online safety information needed by Australian adults is advice on where to report negative online incidents (45%), closely followed by how to use safety and privacy features on devices (43%).[13]

[12] eSafety Commissioner, Protecting Voices at Risk Online, August 2020, https://www.esafety.gov.au/diverse-groups/protecting-voices-risk-online; eSafety Commissioner, Safety by Design, Intersectionality and harm, https://sbd.esafety.gov.au/s/cs/?CS-0021-OH8.

[13] eSafety Commissioner, Building Australian adults' confidence and resilience online, September 2020, https://www.esafety.gov.au/about-us/research/adults-confidence-and-resilience.

eSafety's full suite of research, which explores online attitudes, experiences, harms and solutions, is available online at esafety.gov.au.

Classifying harmful online content

The Online Safety Act creates a framework for the establishment of industry codes and standards. This framework sits within the Online Content Scheme, which provides for the regulation of certain types of harmful online material.

Under the Act, these types of harmful online material are defined as class 1 or class 2 material.

Class 1 and class 2 material

Class 1 and class 2 material are defined under the Act by reference to the National Classification Scheme, a cooperative arrangement between the Australian Government and state and territory governments for the classification of films, publications and computer games.

The National Classification Scheme is implemented through the Classification (Publications, Films and Computer Games) Act 1995 (Cth) (Classification Act) and complementary state and territory enforcement legislation. This complementary legislation sets out how films, publications and computer games can be sold, hired, exhibited and advertised in each state or territory.

The Classification Act establishes the Classification Board and Classification Review Board (the Classification Boards) and sets out their responsibilities and procedures for decision-making. The Classification Act is supplemented by a number of regulations, determinations and legislative instruments, including the National Classification Code (May 2005), Guidelines for the Classification of Publications 2005 (Cth), Guidelines for the Classification of Films 2012 (Cth) and Guidelines for the Classification of Computer Games 2012 (Cth). These provide the principles and criteria for making classification decisions.

Under the National Classification Code, classification decisions are to give effect, as far as possible, to the following principles:

(a) adults should be able to read, hear, see and play what they want;

(b) minors should be protected from material likely to harm or disturb them;

(c) everyone should be protected from exposure to unsolicited material that they find offensive;

(d) the need to take account of community concerns about:

(i) depictions that condone or incite violence, particularly sexual violence; and

(ii) the portrayal of persons in a demeaning manner.

Under the Act, harmful online material is defined by reference to:

- the classification it has received by the Classification Board under the Classification Act (where the material has been classified), or
- eSafety's assessment of the classification the material would likely be given by the Classification Board under the Classification Act (where the material has not been classified).

Classification of content

Online Safety Act	Content	Classification Act/ National Classification Code
Class 1	Film Publication Computer game Any other material*	Refused Classification (RC)
Class 2	Film Any other material (excluding computer games)*	X18+
	Publication	Category 2 restricted
	Film Computer game Any other material*	R18+
	Publication	Category 1 restricted

*Under the Online Safety Act, material that is not a film, computer game or publication is to be classified in a corresponding way to the way in which a film would be classified.

Removal of class 1 and class 2 material

Under the Online Content Scheme, eSafety can issue removal notices for class 1 and class 2 material in certain circumstances. These removal powers are set out in this paper in **3.0 Regulatory responses to harmful online content.**

Currently, the Broadcasting Services Act requires eSafety to apply to the Classification Board for classification of online content before final removal action against Australian-hosted content can occur. The Act changes this to allow eSafety to assess content and take removal action independently of the Classification Board.

Offences relating to class 1 and class 2 material

State and territory enforcement legislation provides for offences in relation to selling, screening, distributing or advertising various categories of classified material (or material that, if classified, would be classified as being in a certain category). State and territory agencies are responsible for enforcement of these laws.

Offences vary significantly across Australian states and territories. While all jurisdictions ban the possession of child sexual exploitation material, the treatment of other types of material is varied. In some states, mere possession of some other forms of Refused Classification (RC) material can be an offence (for example, instructions for manufacturing drugs). However, in most jurisdictions possession of RC material (other than child sexual exploitation material) is only an offence where there is an intent to sell or exhibit the material. The sale or distribution of X18+ films is illegal in most of Australia, though viewing and possession for personal use is not an offence. It is not an offence to possess R18+ material.

Review of Australian classification regulation

On 16 December 2019, the then Minister for Communications, Cyber Safety and the Arts released terms of reference for a review of Australia's classification regulation. This review sought to develop a classification framework that meets community needs and reflects today's digital environment.

The review is currently under consideration by the Australian Government. Due to the connections between eSafety and classification, the Government has been waiting for the finalisation of the online safety legislation before considering possible reforms to classification to allow the continued alignment of the two schemes.

This paper is based on the current classification regime.

Types of material to be considered class 1 and class 2 for industry codes

The National Classification Code and the guidelines for the classification of films, computer games and publications were designed primarily for the assessment of commercially produced material before its release into the community.

They focus on content that may be offensive, more than harmful. They also treat certain categories of behaviour, including specific fetish practices within pornography, as inherently offensive.

The context in which the National Classification Code and classification guidelines were created is very different to the modern online environment. There is now a far greater diversity and volume of content, as well as greater capacity for users to create and distribute content themselves.

Online content involves a mix of written, video and pictorial content.

eSafety views online content as a standalone category of material, separate from films, publications or computer games.

Under the Online Safety Act, material which is not a film, publication or computer game should be classified in a way which corresponds to the way in which films are classified under the National Classification Scheme. Accordingly, eSafety will classify online content in line with the film classification guidelines.

The following table sets out simplified subcategories of class 1 material (1A, 1B and 1C), which are based on, and consistent with, the National Classification Code and film classification guidelines. Class 2 material encompasses two types of material – X18+ material and R18+material. eSafety refers to these subcategories as 2A and 2B in this paper. The simplified subcategories are intended to guide industry. They recognise that some content is more harmful than other content, and industry participants may handle this material in different ways. This is discussed further in **5.0 Preferred codes model.**

eSafety acknowledges that context is important when determining how to classify material and whether it is likely to cause harm. The nature of the material must be considered, including its literary, artistic or educational merit and whether it serves a medical, legal, social or scientific purpose.[14]

Showing crime and violence, for example, may be permissible where the intention is to bear witness to atrocities. Portrayal of drug misuse or addiction, or sexual activity that some may find objectionable, may be permissible where used in public health messaging. Showing, describing or portraying consensual sexual activity may be permissible for adults, whether or not some might find the type of activity offensive, tasteless or immoral.

eSafety considers that industry codes should address class 1 and class 2 material as categorised according to the subcategories in the following table,

[14] See section 11 of the Classification Act.

which seeks to create clarity and certainty for industry and the public by grouping classes of content according to harm. eSafety recognises that classification regimes and laws vary across jurisdictions and accepts that the definitions used by industry participants will never be completely uniform.

Classification and categorisation of class 1 and class 2 material

Class 1 & 2 (Part 9 Online Safety Act)	Subcategories of material to be dealt with by codes	Online material[15]	eSafety harms lens
Class 1 (RC)	**1A**	**CSEM** Child sexual exploitation material.[16] Material that promotes or provides instruction of paedophile activity.	**Harm in production** - Grooming, coercing or threatening a person to produce content - Recording or capturing physical, sexual or psychological abuse; sexual exploitation; or violence to produce online content **Harm in distribution** - Re-traumatisation of victims harmed in the production of content, and violation of their safety, privacy and dignity - Use of material as a recruitment or advocacy tool to threaten, abuse or harm others - Use of material to threaten, harass or abuse people generally, or specific community groups **Harm in consumption** - Feeling disturbed, anxious, upset, scared or traumatised, or becoming desensitised - Normalising the sexualisation of children - Manipulation of beliefs or behaviour, including radicalisation - Contagion or copycat effect, or incitement to violence
		Pro-terror content[17] Material that advocates the doing of a terrorist act (including terrorist manifestos).	
		Extreme crime and violence Material that describes, depicts, expresses or otherwise deals with matters of extreme crime, cruelty or violence (including sexual violence) without justification.* For example, murder, suicide, torture and rape. Material that promotes, incites or instructs in matters of extreme crime or violence.	

[15] These subcategories have been developed to guide industry and are based on, and consistent with, the National Classification Code and film guidelines.
[16] See the Guidelines for the Classification of Films for more information.
[17] This content is deemed to be RC under section 9A of the Classification Act.

<table>
<tr><th>Class 1 & 2 (Part 9 Online Safety Act)</th><th>Subcategories of material to be dealt with by codes</th><th>Online material[15]</th><th>eSafety harms lens</th></tr>
<tr><td rowspan="3">Class 1 (RC)</td><td rowspan="2">1B</td><td>Crime and violence
Material that describes, depicts, expresses or otherwise deals with matters of crime, cruelty or violence without justification.*
Material that promotes, incites or instructs in matters of crime or violence.</td><td rowspan="2">Depending on the circumstances, the production, distribution and consumption of Class 1 - 1B content can involve many of the same harms as Class 1 - 1A. However, this category of content is less well-defined, and more likely to include content that is not prohibited under the laws of other jurisdictions and the terms of service of some platforms.</td></tr>
<tr><td>Drug-related content
Material that describes, depicts, expresses or otherwise deals with matters of drug misuse or addiction without justification.*
Material which includes detailed instruction or promotion of proscribed drug use.</td></tr>
<tr><td>1C</td><td>Online pornography
Material that describes or depicts specific fetish practices or fantasies.[18]</td><td rowspan="3">Harm to children[19]
There is evidence to suggest that exposure to pornography can negatively impact:
- children’s mental health and wellbeing
- their knowledge, attitudes, beliefs and expectations about sex and gender
- their involvement in risky or harmful sexual practices or behaviours.
Content relating to violence, drug use, language and themes can also be harmful to children – particularly where the content is detailed, prolonged, realistic and/or interactive.</td></tr>
<tr><td>Class 2 (X18+)</td><td>2A</td><td>Online pornography
Other sexually explicit material that depicts actual (not simulated) sex between consenting adults.</td></tr>
<tr><td>Class 2 (R18+)</td><td>2B</td><td>Online pornography[20]
Material which includes realistically simulated sexual activity between adults.
Material which includes high-impact[21] nudity.</td></tr>
</table>

[18] See the Guidelines for the Classification of Films for more information.
[19] eSafety acknowledges that pornography and other high impact content also has the potential to cause harm to adults in some circumstances.
[20] Context is required to determine whether this content is pornographic or ‘other high impact content’
[21] Impact may be higher where content is detailed, accentuated, or uses special effects, prolonged, repeated frequently, realistic or encourages interactivity.

Class 1 & 2 (Part 9 Online Safety Act)	Subcategories of material to be dealt with by codes	Online material[15]	eSafety harms lens
		Other high impact content Material which includes high-impact[22] sex, nudity, violence, drug use, language and themes. 'Themes' includes social Issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism.	

* The nature of the material must be considered, including its literary, artistic, or educational merit and whether it serves a medical, legal, social or scientific purpose. See section 11 of the Classification Act.

Industry responses to harmful online content

eSafety acknowledges and welcomes the online industry's current efforts to address the risks and harms associated with class 1 and class 2 material.

The following table provides examples of some of these key initiatives.

Initiatives	Examples
Providing information and advice	- Online safety resources that explain online risks and harms and promote the use of safety tools.
Developing and publishing policies	- Terms of use, acceptable use policies, terms of service, community standards and/or rules which set out what is and is not allowed on the service.
Implementing systems to prevent, detect and address harmful material and activity	- Age assurance mechanisms to establish or predict the age (or age range) of the user. - Hashing technology to convert images into a unique signature which can be used to detect, notify and remove child sexual exploitation material and pro-terror content. - Artificial intelligence, machine learning and deep learning classifiers that seek to identify and prevent upload or flag for review harmful text, images, videos, audio, live-streams and/or grooming or predatory behaviour. - Moderation teams, practices and procedures to detect, review, notify and take action against harmful content and activity, including through warning account-holders, suspending accounts, removing content, placing it

[22] Impact may be higher where content is detailed, accentuated, or uses special effects, prolonged, repeated frequently, realistic or encourages interactivity.

	behind content warnings or restricted access systems, and de-monetising or deindexing it from searches or recommendations.
Providing tools to empower users to manage their online experience	- Mechanisms for users to flag, report or make complaints. about harmful material or activity they encounter. - Safety and privacy features and settings. - Parental controls and filtering tools. - Community-based moderation.
Collaborating with others	- Sharing hash sets or URLs known to be providing access to child sexual exploitation material or pro-terror content. - Open sourcing detection and moderation technologies. - Supporting research and innovation. - Contributing to cross-sector online safety groups and initiatives.

Challenges to addressing harmful online content

Despite industry efforts to address online content risks and harms, significant challenges remain.

Firstly, there are jurisdictional challenges that arise when working across legislative borders and across non- or under-regulated environments.

Over 99% of child sexual exploitation material (CSEM) investigated by eSafety is hosted outside of Australia, so removal action is pursued through international relationships with law enforcement collectives such as INHOPE. While legal definitions vary between countries, some international baselines have been agreed to facilitate cross-border efforts to combat CSEM. However, other types of harmful content are defined and treated inconsistently across countries.

Similarly, the definition and treatment of harmful content often varies across services. This is particularly the case for lower end class 1 and class 2 material, especially online pornography. For example, some services choose to ban all content depicting sexual activity, while others require it to be labelled and placed behind a content warning or restricted access system, and still others provide easy access and actively promote the content. For some, providing paid access to this material is their business model.

For the higher end of class 1 material involving material such as CSEM or pro-terrorist content, a significant number of industry participants are working to prevent their services and devices from being used to produce, access, store or distribute this material. Industry participants who already have robust measures in place will be able to draw on existing efforts and initiatives to demonstrate compliance with the new industry codes.

However, this is not true of all industry participants. For example, eSafety has encountered services that lack effective content reporting mechanisms, fail to find and address overt CSEM and are non-responsive to removal requests. These services may deny ownership or connection to harmful content, refuse to acknowledge its presence on their systems or show a lack of will to remove it.

Other services may attempt to address harmful content in good faith, but may have limited resources, capacity and capability to do so. This is especially the case for start-ups and small companies. The need to balance risks and

resources, while still meeting appropriate standards of online safety, is discussed in more detail in **4.0 eSafety's positions on codes development.**

Moreover, all efforts to address harmful content are occurring in a rapidly evolving online environment where perpetrators are constantly adapting to evade detection. Removed content can easily be uploaded again as a different file, bad file hosts can change servers, and images and videos can be edited to avoid detection software. The use of decentralised platforms, encrypted services and identity shielding tactics, such as virtual private networks that mask the user's location and device details, create additional challenges that need to be met with innovative solutions. To be effective, these solutions should be focused on proactive and systemic change, rather than only the removal of individual items of content after they have already been shared. This is particularly important for services at risk of being used to host or provide access to higher end class 1 material, such as CSEM and pro-terror content.

eSafety expects that the industry codes will extend beyond ensuring reactive removal to support proactive measures to prevent, disrupt and alleviate harm. The codes should require all industry participants to consider their role in the digital ecosystem and the risk profile of their services and devices.

The application of the codes to a wide spectrum of industry sections should provide a foundation for collaboration between industry participants. In particular, larger and more mature services should be encouraged to assist in capacity-building of smaller and newer services.

3.0 Regulatory responses to harmful online content

International regulatory responses

The online safety landscape is undergoing significant change. Governments across the world are increasingly recognising the need for industry to act proactively to address online harms. In response, several countries have introduced or are considering online safety regulatory initiatives, including industry codes.

eSafety has explored the approaches taken by other governments and international regulators in order to inform the approach in Australia. eSafety is also conscious that harmonising, where possible, online safety codes and governance arrangements across jurisdictions is likely to help prevent fragmentation that could hinder the effectiveness of global and local online safety efforts.

The following table shows some key international examples.[23] A more detailed analysis of these examples is set out at Appendix C.

	United Kingdom (UK)	European Union (EU)	Germany	Ireland
Code	Interim Code of practice on terrorist content and activity online (2020) Interim Code of practice on online child sexual exploitation and abuse (2020)	Code of conduct on countering illegal hate speech online (2016)	Safety by design requirement to implement pre-emptive protection measures, under the Youth Protection Act, Section 24a (2021)	Online safety codes under the Online Safety and Media Regulation Bill 2020
Compliance	Voluntary	Voluntary	Mandatory	Will be mandatory
Type of content	User-generated online terrorist content or activity.	Hate speech, including racist and xenophobic content and	The requirements target content that is harmful	The online safety codes will broadly cover

[23] eSafety has limited its review in this paper to codes intended to address harmful content and activity online. However, eSafety is aware of a number of international codes which seek to address closely related issues. For example, the UK Information Commissioner's Age appropriate design code protects children's data online. While the Age appropriate design code focuses on privacy, it has obvious impacts on and synergies with safety. eSafety conducts broad scanning and engagement to ensure eSafety is aware of, and to the extent possible, harmonised with, these efforts.

	United Kingdom (UK)	European Union (EU)	Germany	Ireland
	Child sexual exploitation and abuse.	terrorist propaganda.	to minors or impairs their development.	harmful online content.
Scope of services	Terrorist and violent extremist content: companies that supply services or tools which allow, enable or facilitate users to share or discover user-generated content or to interact with each other online. Child sexual exploitation and abuse: All companies, including small- and medium-sized enterprises, noting that not all principles apply to every company.	Signatories include major platforms such as Facebook, Microsoft, Twitter, YouTube, Instagram, Snapchat, Google+, Dailymotion, TikTok and LinkedIn.	Service providers that store or provide third-party information for/to other users for profit. Exemptions include services not directed at or typically used by minors.	The Commissioner decides which codes apply to which online services, including to whole categories. A wide range of services may be designated by the Commissioner.
Compliance and enforcement	The Interim Codes and all the principles contained within them are voluntary and non-binding. Each principle includes examples of good practice, but no mechanism for enforcement.	Compliance is evaluated through regular monitoring in collaboration with a network of organisations located in the different EU countries. Using a commonly agreed method-ology, these organisations test how participants are implementing the commit-ments in the Code.	The largest service providers must implement guidelines that are agreed by an industry body and approved by the Federal Agency for the Protection of Children and Young People Within Media, while industry bodies must determine whether the guideline apply to a particular service. Compliance is reviewed by a dedicated youth protection organisation. Enforcement measures	The Commissioner will evaluate compliance through audits and periodic reports from services. Services will need to follow rules on periodic reporting. The Commissioner can enforce rules through measures including compliance and warning notices, statutory investigations, financial sanctions, and the termination or suspension of

	United Kingdom (UK)	European Union (EU)	Germany	Ireland
			include issuing an order to implement measures, and a fine of up to EUR 50 million.	contracts of licensed services. The Commissioner will take a proportionate and balanced risk-based approach to enforcing its Codes.
Status	Interim Codes have been released. The UK communications regulator is expected to develop binding codes under the proposed Online Safety Bill, which is before UK Parliament.	In force.	In force.	Intention to publish online safety codes has been outlined in the Online Safety Media and Regulation Bill, which is currently before Irish Parliament.

Each jurisdiction has approached codes differently. Australia's approach will be unique, with binding codes applicable to all participants within the eight industry sections outlined in the Online Safety Act. The approaches taken by other jurisdictions, however, provide some valuable ideas that eSafety recommends for industry codes in Australia.

These include:

- addressing online safety issues by themes and topics
- using language and terms that industry and consumers readily understand
- focusing on principles and outcomes that highlight good practice responses
- taking risk-based approaches
- considering the functionality of services and how it impacts harms
- recognising the unique status of children and the importance of keeping them safe from harm online.

Industry associations developing codes under the Act should remain aware of international approaches when developing codes.

Australia's regulatory framework

Online Content Scheme

As outlined previously, the Online Safety Act will become operational on 23 January 2022.

Part 9 of the Act establishes a more modern and effective Online Content Scheme which:

- sets out the eSafety Commissioner's removal powers with respect to class 1 and class 2 material, including expanded powers to issue removal notices for class 1 material, no matter where that content is hosted. eSafety may also require certain class 2 material to either be removed or placed behind a restricted access system, which limits the exposure of children to age-inappropriate materials
- gives the eSafety Commissioner the power to require search engines and app stores to remove access to websites or apps in certain circumstances where removal notices for class 1 material have not been effective
- sets out the regime for new industry codes and standards
- gives the eSafety Commissioner the power to make service provider determinations in certain circumstances
- gives the eSafety Commissioner the power to apply to the Federal Court for an order to stop the provision of a particular social media service, relevant electronic service designated internet service or internet carriage service in certain circumstances.

eSafety can investigate certain matters in respect of class 1 and class 2 material on its own initiative or in response to complaints from members of the public. The following table sets out when a complaint can be made to eSafety about class 1 or class 2 material and eSafety's removal powers.

Class	Complaints to eSafety	eSafety removal power*
Class 1	Yes, where a person has reason to believe that end-users in Australia can access material provided on a social media service, relevant electronic service or designated internet service	The material must be removed no matter what country it is hosted in or provided from
Class 2 (X18+)		The material must be removed if it is hosted in or provided from Australia
Class 2 (R18+)	Yes, where a person has reason to believe that end-users in Australia can access material provided on a social media service, relevant electronic service or designated internet service and the material is not subject to a restricted access system	The material must be placed behind a restricted access system or removed, if it is hosted in or provided from Australia

* eSafety may issue removal notices to social media service providers, relevant electronic service providers, designated internet service providers and hosting service providers.

A complaint can also be made to the eSafety Commissioner if a person believes that someone has breached an industry code or standard or a rule set out in a service provider determination. eSafety can investigate these matters on its own initiative or in response to complaints. The Act affords the eSafety Commissioner the power to conduct investigations as the Commissioner sees fit.

Overview of industry codes

Modern, effective industry codes will be integral to the operation of the new Online Content Scheme.

The purpose of the codes is to create a whole of industry response to class 1 and class 2 material. Further, the codes will be part of a broader framework for improving and maintaining online safety, complementing eSafety's other measures and initiatives that help Australians have safer, more positive experiences online.

The codes are intended to apply to the participants of eight key sections of the online industry[24] in respect of their online activities, as set out in the following table. To the extent that any services offer encryption or private messaging, the codes will apply to these services.

[24] The Online Safety Act provides that the Minister may, by legislative instrument, make rules (legislative rules) specifying exempt services. At present, there is no intention to make legislative rules to exempt services from the codes.

Section of the online industry		Scope
Social media services	Providers of social media services, so far as those services are provided to end-users in Australia	All providers of social media services that can be accessed by end-users in Australia, including: - social networks - media sharing networks - discussion forums - consumer review networks
Relevant electronic services	Providers of relevant electronic services, so far as those services are provided to end-users in Australia	All providers of relevant electronic services that can be accessed by end-users in Australia, including: - email services - instant messaging services - SMS and MMS services - chat services - online games where end-users can play against each other[25] - online dating services
Designated Internet services	Providers of designated internet services, so far as those services are provided to end-users in Australia	All providers of designated internet services, such as websites[26] that that can be accessed by end-users in Australia (unless a service is otherwise considered a social media service or a relevant electronic service)
Search engine services	Providers of internet search engine services, so far as those services are provided to end-users in Australia	All providers of search engine services that can be accessed by end-users in Australia. Search engine services are software-based services designed to collect and rank information on the World Wide Web (WWW) in response to user queries. Search engine services exclude search functionality within platforms where content or information can only be surfaced from that which has been generated/uploaded/created within the platform itself and not from the WWW more broadly
App distribution services	Providers of app distribution services, so far as those services are provided to end-users in Australia	All providers of app distribution services that can be accessed by end-users in Australia. App distribution services exclude: - links to an app distribution service

[25] The codes will not apply to game content which has been classified in Australia. However, the codes will apply to content imported into a game environment via the game's interactive tools which is separate to the game itself and which is likely to be classified as class 1 or class 2 material. The codes will also apply to game content which has been recorded and posted elsewhere on the internet.

[26] Not all websites present a risk in respect of class 1 and class 2 material. How the codes are to apply to websites is discussed further at **4.0 eSafety's positions on codes development.**

Section of the online industry		Scope
		- download of apps from third party websites
Hosting Services	Providers of hosting services, so far as those services host material in Australia	All hosting services providers which host stored material in Australia (for example, where a service has data centres located in Australia)
Internet Carriage services	Providers of internet carriage services, so far as those services are provided to customers in Australia	All internet service providers who provide internet access to customers in Australia
Manufacturing, supplying, maintaining or installing equipment	Persons who manufacture, supply, maintain or install equipment for use by end users in Australia of a: - social media service in connection with the service - relevant electronic service in connection with the service - designated internet service in connection with the service - internet carriage service in connection with the service	All persons who manufacture, supply, maintain or install equipment used by end-users in Australia to: - browse the internet – including mobile phones, laptops, tablets, internet-enabled devices (such as smart TVs and gaming consoles) and immersive technologies (such as virtual reality headsets) - connect to the internet, such as wi-fi routers. This section of the online industry includes manufacturers of these devices, as well as businesses and retail outlets that install, sell and/or repair/maintain such devices

Part 9, Division 7 of the Act outlines the principles for the development of industry codes:

- Section 137(2) requires the eSafety Commissioner to make reasonable efforts to ensure that, for each section of the online industry, either a code is registered within 6 months of the commencement of the Act, or an industry standard is registered within 12 months of the commencement of the Act.
- Section 138(3) outlines examples of matters that may be dealt with by industry codes and/or standards.
- Section 140 outlines the process and requirements for the eSafety Commissioner to register an industry code.
- Section 141 outlines a process by which the eSafety Commissioner may request that an industry body or association develop a code.
- Section 141A enables the eSafety Commissioner, in circumstances where no body or association represents a section of the online industry, to publish a notice stating that if an industry or association were to come into effect,

eSafety would likely request that body or association develop a code under section 141.

- Section 142 clarifies that a code will be replaced rather than varied where changes are made.

These matters are discussed further in **4.0 eSafety's positions on codes development** and **6.0 Registration process.**

Industry codes are mandatory and enforceable. Under the Act:

- Section 143 outlines the eSafety Commissioner's power to direct compliance with a code. Failure to comply with a direction may attract a civil penalty of 500 penalty units (up to $111,000 for individuals and up to $555,000 for companies).
- Section 144 provides the eSafety Commissioner with the power to issue a formal warning for breach of an industry code.

Compliance and enforcement of industry codes is discussed further in **4.0 eSafety's positions on codes development.**

Overview of industry standards

The circumstances and conditions under which the eSafety Commissioner may determine an industry standard are outlined in Section 145.

In summary, these are when:

- the eSafety Commissioner has made a request for code development under section 141 that is not complied with or one of a number of other conditions are satisfied, including that the draft code does not contain appropriate community safeguards for matters specified in the request
- the eSafety Commissioner has published a notice under 141A and no industry body or association comes into existence within the period specified, or
- the eSafety Commissioner is satisfied that a code that has been registered for at least 180 days is deficient and the Commissioner is satisfied that notified deficiencies have not been adequately addressed within a specified period.

In addition, the eSafety Commissioner must not determine a standard unless satisfied that it is necessary or convenient to provide appropriate community safeguards or otherwise adequately regulate participants in an online industry section.

Broader online safety framework

In addition to industry codes and/or standards and the other provisions of the Online Content Scheme, matters relating to class 1 and class 2 material may also be addressed by the Basic Online Safety Expectations (BOSE), a regime under the Act that sets out the Government's expectations for online safety.

The regulation of class 1 and class 2 material can be understood as different layers of a pyramid, with formal regulatory powers underpinned by co-regulatory measures. These elements are built on a solid foundation of wide-ranging online safety expectations about online content and activity laid out in the BOSE. All of these elements work together to achieve the same outcome of keeping Australians safe online.

Regulation of
class 1 and class 2 material

Formal regulation

Removal and remedial action

- Industry standards
- Class 1 and class 2 removal notices
- Class 2 remedial notices (to place content behind a restricted access system)
- Link deletion and app removal notices for class 1 material

Co-regulation

Industry codes

Mandatory codes developed by industry to address online harms relating to class 1 and class 2 material.
Apply to 8 industry sections.
eSafety can direct industry participants to comply.

Objective 1:
Industry participants will take proactive steps to create and maintain a safe online environment

Objective 2:
Industry participants will empower people to manage access and exposure to class 1 and class 2 material

Objective 3:
Industry participants will strengthen transparency of, and accountability for, class 1 and class 2 material

Basic Online Safety Expectations

Australian Government's expectations for online safety established by Ministerial instrument.
Apply to social media services, relevant electronic services and designated internet services only.
eSafety can require these services to report on how they are meeting the expectations.

Division 2:
Expectations regarding safe use

Division 3:
Expectations regarding certain material and activity

Division 4:
Expectations regarding reports and complaints

Division 5:
Expectations regarding making certain information accessible

Division 6:
Expectations regarding record keeping

Division 7:
Expectations regarding dealings with the Commissioner

Note: the BOSE divisions are set out in the draft instrument but are not final

Safety by Design

Voluntary guidance material and tools developed by eSafety in collaboration with industry to enable companies to lift their safety capabilities to better meet Government's expectations.
Applies to all services and products.

Principle 1:
Service provider responsibility

Principle 2:
User empowerment and autonomy

Principle 3:
Transparency and accountability

Layer 1: Basic Online Safety Expectations

On 8 August 2021, the Department of Infrastructure, Transport, Regional Development and Communications commenced consultation on the draft Online Safety (Basic Online Safety Expectations) Determination 2021. The BOSE are the Government's expectations for certain sections of the online industry with respect to all aspects of online safety (both content and activity) and not just the management of harmful online content categorised as class 1 or class 2 material. They apply to social media services, designated internet services and relevant electronic services. These services will be held to account by the eSafety Commissioner, who can require them to report on their compliance with the BOSE.

The draft Determination includes the Government's core expectations for the online industry, which are specified in the Act, as well as a number of additional expectations.

There is some potential for overlap between the BOSE and the codes. For example, core expectations in the BOSE include:

- the provider of a service will take reasonable steps to minimise the extent to which class 1 material is provided on the service
- the provider of a service will take reasonable steps to ensure that technological or other measures are in effect to prevent access by children to class 2 material provided on the service.

However, the BOSE and the codes serve different regulatory purposes. The purpose of the BOSE is to place greater responsibility on service providers to ensure they provide safer services to Australian end-users. The codes are directed to ensuring that class 1 and class 2 material is prevented, or limited, on services accessible to Australian end-users.

Layer 2: Co-regulation via industry codes

Given the potential for overlap between the BOSE and the industry codes, it is important they are aligned. Accordingly, in crafting objectives and outcomes for the codes in **5.0 Preferred codes model**, eSafety has built upon the foundation provided by the core expectations set out in the Act. Compliance with the codes may assist social media services, designated internet services and relevant electronic services in demonstrating that they are meeting some expectations under the BOSE, and vice versa. For example, the codes may

require some services to take measures to detect and prevent, or limit, some categories of class 1 material. Taking those measures would assist services both to comply with the codes and also potentially some of the core expectations in the BOSE which relate to minimising the provision of class 1 material.

Layer 3: Formal regulation

Formal regulation under the Act sits at the sharp end of the pyramid because it includes specific tools that should only need to be used infrequently, if industry is meeting its obligations under the codes and BOSE. This is where industry standards would sit, if they are necessary. The tools include removal notices, link deletion notices and app removal notices.

Remedial notices also sit within the formal regulation layer. Under the Act, the eSafety Commissioner can issue a remedial notice requiring content provided from Australia or hosted in Australia which has been (or would likely be) classified R18+ or category 1 restricted under the National Classification Code to be either removed or placed behind a restricted access system (RAS). A RAS is the means by which children are protected from exposure to this age-inappropriate material. On 16 August 2021, eSafety released a discussion paper on development of the new RAS declaration. Consultation closed on 12 September 2021. Public consultation on the RAS instrument is scheduled to take place in October 2021.

Other relevant eSafety initiatives

Age verification roadmap

On 1 June 2021, the Australian Government tabled its response to the 'Protecting the age of innocence' report by the House of Representatives Standing Committee on Social Policy and Legal Affairs. The response indicated support for recommendation 3, which called for the eSafety Commissioner to lead development of an implementation roadmap for a mandatory age verification regime that limits access to online pornography.

Given the close relationship between the age verification roadmap and the RAS, eSafety began these processes concurrently. Accordingly, on 16 August 2021, eSafety also commenced the consultation processes for age verification. The report is due to Government in December 2022.

eSafety's recommendations and options to Government will be informed by the advice of the Australian community, including the civil, academic, not-for-profit and industry sectors.

eSafety is gathering research, intelligence and information on the range and effectiveness of technologies and systems that assist in minimising children's exposure to online pornography, as well as other measures that could support an age verification framework.

These include:

- identifying good practice in educating young people on pornography and respectful relationships
- educating parents and carers about discussing pornography and related content (as well as reporting it)
- running campaigns to educate the public on the intent and functionality of age verification technologies
- providing information on other effective technological measures, such as parental controls and device-level filtering, which help prevent children's access to pornography.

Codes content which is covered by the Restricted Access System and the Age Verification Roadmap

<table>
<tr><th>eSafety subcategories of material</th><th>Material</th><th>Restricted Access System</th><th>Age Verification Roadmap</th></tr>
<tr><td rowspan="3">1A</td><td>CSEM</td><td rowspan="6">Does not apply to this material</td><td rowspan="3">Does not apply to this material</td></tr>
<tr><td>Pro-terror content</td></tr>
<tr><td>Extreme crime and violence</td></tr>
<tr><td rowspan="2">1B</td><td>Crime and violence</td></tr>
<tr><td>Drug-related content</td></tr>
<tr><td>1C</td><td>Online pornography</td><td rowspan="3">Applies to this material</td></tr>
<tr><td>2A</td><td>Online pornography</td></tr>
<tr><td rowspan="2">2B</td><td>Online pornography</td><td rowspan="2">Applies to this material</td></tr>
<tr><td>Other high impact content</td><td>Does not apply to this material</td></tr>
</table>

Safety by Design

Separately, industry is encouraged to adopt the Safety by Design principles set out by eSafety following wide-ranging consultation.

The Safety by Design initiative puts user safety and rights at the centre of the design, development and deployment of online devices and services. eSafety has provided the online industry with a set of voluntary guidance materials and interactive assessment tools that help companies to assess online safety risk for users and improve safety protections and capabilities. Applying the Safety by Design principles will help industry participants meet and exceed the Government's expectations.

The tools can be used by a broad range of companies across the digital ecosystem, but are targeted towards services and devices that allow users to interact or engage with others, share or post content, or provide an internet-connected interactive experience. The tools have been built to be interactive, educative and informative, guiding participants through assessment questions and tangible examples of good practice.

The tools are divided into two categories – one for early-stage technology companies (with 0 to 49 employees worldwide) and another for mid-tier companies (with 50 to 249 employees worldwide) and enterprise companies (with 250+ employees worldwide). This ensures they can be used by companies at every stage of the design, development and deployment process and tailored to their level of risk and resourcing.

Other relevant Australian codes

eSafety has also engaged with other Australian regulators with similar or interconnected purposes and functions. This includes the Australian Communications and Media Authority (ACMA), the Office of the Australian Information Commissioner (OAIC), the Australian Competition and Consumer Commission (ACCC) and the Department of Home Affairs. Each has administered, or been tasked with administering, industry code processes relating to digital platforms or digital devices.

Several of these code processes arose from recommendations of the ACCC's Digital Platform Inquiry, which the Government supported.

Of note:

- The ACMA was tasked with overseeing a voluntary code addressing disinformation and misinformation. The final code, the Australian Code of Practice on Disinformation and Misinformation, was published on 22 February 2021.
- The ACCC was tasked with developing a mandatory code of conduct to address bargaining power imbalances between Australian news media businesses and digital platforms, specifically Google and Facebook. The final legislation passed both Houses of Parliament on 25 February 2021.
- The OAIC was tasked with developing a mandatory code addressing the privacy and data practices of digital platforms. This process is yet to be finalised.

In addition, Home Affairs oversaw the development of the Voluntary Code of Practice: Securing the Internet of Things for Consumers, which was released on 3 September 2020. Based on research findings that reviewed the first six month of the code's operation, Home Affairs sought feedback on the merits of a mandatory cyber security standard for smart devices in Australia and/or cyber security labelling.

Where possible, eSafety is committed to collaborating and coordinating with other Australian regulatory agencies to promote complementary code approaches that ease compliance for industry participants.

4.0 eSafety's positions on codes development

This chapter sets out 11 positions for achieving robust codes that provide appropriate community safeguards. When eSafety receives a code for registration, it will consider the extent to which the code aligns with these 11 positions.

eSafety acknowledges that the Online Safety Act gives industry associations discretion as to how to structure the codes and eSafety will consider alternative approaches. However, it expects that industry will explain how any deviations achieve the policy intent of these positions, including the provision of appropriate community safeguards.

Substance of the codes

Position 1: The codes will address the issues of access, exposure and distribution that are related to class 1 and class 2 material.

Industry codes need only address issues that are related to class 1 and class 2 material.

However, the risks and harms associated with class 1 and class 2 material are not limited to access and exposure. There are also risks and harms associated with the distribution of the material, including live streaming. eSafety also expects the codes to address the online storage of class 1 – 1A material.

A wide range of harms are associated with child sexual exploitation material that is generated by live streaming or recording a child being abused. This abuse results in grave harm to the child, regardless of whether any lasting content is produced. However, where the material is saved and stored online, the possibility that it may be shared and viewed by others can cause further serious distress to the child.

eSafety expects that the codes will address the full range of risks and harms relating to class 1 and 2 material. This ought to extend, where reasonably possible, to the facilitation of the production of class 1 - 1A material. For example, contact or messaging involving the grooming of a child to facilitate the production of child sexual abuse material. The codes should prompt

industry participants to consider their role in the digital ecosystem in order to implement reasonable and proportionate measures to address this risk.

eSafety also expects that services with capacity to analyse their user base will consider and take reasonable steps to address the needs of at-risk groups. This helps ensure online safety interventions adequately address multidimensional risks and do not make them worse for disadvantaged or marginalised groups.

Position 2: The application of the codes will not be limited to services provided from Australia.

Some of the eSafety Commissioner's removal powers under the Online Content Scheme only apply to services 'provided from Australia'. For example, services who have a registered office or carry-on business in Australia.

This limitation does not apply to industry codes and standards.

With the exception of hosting service providers, the codes apply to all online service providers, regardless of where they are located, so far as their services are provided to end-users or customers in Australia. The codes also apply to manufacturers, suppliers, maintainers or installers of devices, where those devices are for use by Australian end-user to access online services.

The codes apply to hosting service providers to the extent they host content in Australia.

A narrower application would result in a failure by industry to address large volumes of content available to Australians, such as online pornography provided from companies based overseas.

The codes will complement the Commissioner's removal powers with respect to class 1 and class 2 material, which apply in more limited circumstances and which can be used where the industry has otherwise failed to address harmful content under the codes.

Design of the codes

Position 3: Industry associations will develop a set of common drafting principles to inform codes development.

A common and agreed set of drafting principles should underpin the development of industry codes. This will ensure codes are developed in a consistent, coordinated and streamlined manner.

The drafting principles should incorporate the following elements.

Guiding principles

- **Consistent:** The codes contain a consistent approach to the regulation of class 1 and class 2 material across industry sections. If there are multiple codes, the codes do not contain overlapping or inconsistent measures which could be contradictory or confusing.
- **Clear:** The codes are drafted in plain language, so that they are understood by the public and industry participants. The scope and application of each code is clearly defined.
- **Meaningful:** The codes improve online safety standards for Australians, by requiring industry participants to proactively address safety issues, collaborate with other services, empower users, publish policies and provide safety information and advice to the public.
- **Implementable:** Industry participants understand what is required of them for codes compliance and have the flexibility to innovate and adapt their compliance measures over time.
- **Measurable:** Compliance with the codes can be assessed and enforcement action taken if needed.
- **Proportionate:** The codes acknowledge the diversity of size, maturity, capacity and capability of industry participants, while requiring all participants do their part to improve user safety.
- **Respectful of rights:** The codes consider the protection and promotion of human rights online, including the right to freedom of expression and access to information and privacy, while also the freedom from violence, abuse and discrimination. The codes also consider the best interests of children and the rights of victims, whose images or other details have been shared online.

- **Balance safety, privacy and security:** The codes take a balanced approach to safety, privacy and security.

Position 4: The codes will adopt an outcomes- and risk-based regulatory approach, supported by clear compliance measures which apply to industry participants whose services or devices present the greatest risk in respect of class 1 and class 2 material.

eSafety considers that industry associations should adopt a regulatory approach that is both outcomes-based and risk-based when developing the codes.

An outcomes-based approach provides industry participants with a common set of objectives and outcomes, while granting the flexibility to implement measures to meet those objectives and outcomes that are most suited to their business models and technologies. These measures should be reasonable and proportionate, based on an assessment of the risk an industry participant's services and/or devices present in respect of class 1 and class 2 material.

However, to ensure high-risk industry participants can be held to account, eSafety considers that the outcomes-based model should be supported by clearly defined minimum compliance measures for each outcome.

An overview of an outcomes-based regulatory approach is outlined below.

An outcomes-based regulatory approach

An outcomes-based regulatory approach involves establishing an agreed set of clear and measurable outcomes that describe what a code is seeking to achieve, without prescribing how those outcomes are to be achieved. This is in direct contrast with prescriptive, rules-based regulatory approaches, which apply a 'one size fits all' model to all regulated participants.

The primary distinguishing features of an outcomes-based approach are:

1. **Regulation is drafted with high-level outcomes or objectives that must be met.**
2. **Industry participants develop their own systems to achieve the outcomes specified in the regulation.**
3. **Industry participants are required to demonstrate delivery of these outcomes to the regulator, with enforcement and compliance measures in place should a failure to achieve an outcome occur.**

Unlike more rigid, rules-based approaches, outcomes-based regulation is well suited to complex and dynamic regulatory environments, such as the online industry.

This approach:

1. **Gives participants greater flexibility to develop and implement systems to comply with regulated outcomes, including the adoption of emerging technologies.**
2. **Enhances the responsiveness of the regulatory environment to adapt to new and novel harms.**
3. **Minimises the risk that the regulatory environment will become outdated or unworkable.**

Outcomes-based regulation can also comfortably co-exist with other regulatory approaches, to achieve a balance between outcomes and rules.

The success of an outcomes-based approach is dependent on several key conditions, including cooperation and support by industry participants and clearly defined outcomes. A successful outcomes-based approach also requires:

- **robust compliance and enforcement mechanisms**
- **a transparent program of evaluation, review and reporting to demonstrate compliance.**

eSafety considers that there are clear benefits of an outcomes-based regulatory approach, given the breadth of online activities covered by the eight sections of the online industry to which the codes apply, as well as the potential for emergence of unforeseen online harms.

eSafety has developed a model which articulates its preferred objectives and outcomes for the codes. This model is set out in **5.0 Preferred codes model.**

The outcomes-based model should apply to all codes to ensure a consistent regulatory approach. However, given the diverse nature of the online activities covered by the codes, not all of the outcomes may apply to every industry section. For example, an outcome aimed at the proactive detection and removal of harmful material may not apply to manufacturers, suppliers, maintainers and installers of devices.

A risk-based approach

The measures adopted by industry participants to meet the objectives and outcomes of the codes should be reasonable and proportionate, based on an

assessment of the risk an industry participant's services and devices present in respect of class 1 and class 2 material. Each industry participant should consider the likelihood that its:

- services are used to store class 1 - 1A material;
- services and/or devices result in users (particularly children) being exposed to class 1 and class 2 material;
- services and/or devices are used to access class 1 and class 2 material; and
- services and/or devices are used to distribute class 1 and class 2 material.

Applying a risk-based approach is entirely consistent with the application of outcomes-based regulation and is particularly appropriate given the sections of the online industry cover a broad range of services and devices which prevent varying degrees of risk. eSafety will not limit the scope of any sections of the online industry by creating exclusions or carve-outs for industry participants who provide any particular types of services or devices.

Challenges of an outcomes-based approach

There can be some risks associated with an outcomes-based regulatory approach, including:

- potential tension between the regulator and the regulated participants' understanding of what a satisfactory regulatory outcome looks like in practice
- potential lack of certainty for regulated participants about what actions they need to take to comply with regulated outcomes.

Small and medium sized companies, in particular, may experience challenges with an outcomes-based code, as they are likely to have less capacity and fewer resources than large companies to develop and implement compliance measures.

However, the devices and online services of industry participants of all sizes can be used to access and distribute harmful online material, such as CSEM.

A combination of outcomes and rules

To address these concerns, eSafety encourages the online industry to consider an outcomes- and risk-based regulation model supported by clearly defined minimum compliance measures for each outcome. These measures should apply to industry participants whose services and devices present the greatest risk in respect of class 1 and class 2 material.

The intent of this approach is to ensure that the codes contain clear expectations for services and devices which carry the most risk.

High risk industry participants are also encouraged to continue to do as much as they can to minimise online harms, beyond the minimum compliance measures.

Minimum compliance measures should be applied to industry participants of all sizes, from large companies to start-ups. This may mean that the codes require different minimum compliance measures for different industry participants, based on the size, maturity, capacity and capability of an industry participant, as well as risk level. This could include whether industry participants identify as early-stage companies (0-49 employees worldwide), mid-tier companies (50-249 employees worldwide) or enterprise companies (250+ employees worldwide).

While other jurisdictions have segmented legislative obligations according to company size, exempting small companies from certain requirements,[27] eSafety considers that this type of blanket approach would not be appropriate in the context of the codes developed under the Online Safety Act. It is well recognised that bad actors exploit smaller platforms for CSEM and pro-terror content. Accordingly, codes intended to address risks and harms associated with this type of class 1 material would fail to achieve their goal if small services were to be exempted. Instead, eSafety encourages industry to develop a range of minimum compliance measures which leverage existing good practice and take a balanced approach to risks and capacity constraints.

Determining risk profile

eSafety encourages industry to consider how the risk profile of each industry participant is best established. eSafety recommends that industry develop a set of agreed risk categories (for example, high, medium and low) and a set of objective criteria for each industry section, so that each industry participant can be easily identified as being, for example, high risk or low risk. Some industry sections (for example, social media services and relevant electronic services) may be able to use the same risk matrix.

It may also be that the nature of the industry section means that all industry participants share similar risk profiles. For example, providers of internet

[27] See, for example, the European Union Digital Services Act.

carriage services may all present the same level of risk with respect to class 1 and class 2 material. If this is the case, then the codes should set out minimum compliance measures for all participants (but may, for example, provide different measures based on the size of the industry participant).

A number of factors are relevant to assessing the risk level of services and devices as set out in the following tables. This list is not an exhaustive list of relevant factors but is intended to provide industry with a guide.

Services

Functionality of a service	The functionality of a service is directly relevant to: - the role the service plays in facilitating access or exposure to, and/or distribution of, class 1 and class 2 material, as well as online storage of class 1-1A material; and - the measures available to the service to prevent or limit access to class 1 and class 2 material. For example, the following functionalities will likely increase a service's risk profile: - allowing end-users in Australia to link to, or interact with, other end-users (for example, via voice/video calls, live streaming, text messaging) - allowing end-users in Australia to generate, store, post or share material (for example, via content upload and file sharing)
Purpose of a service	The primary purpose of a service will be a relevant consideration in assessing risk. For example, services: - developed for government use - designed by schools or universities for educational or research purposes, will likely have a lower risk profile than services designed purely for social interaction and networking. Conversely, a service developed primarily to provide access to online pornography would have a high risk profile. Industry participants should remain aware that secondary purposes may also carry risk.
Nature of the user base	The user base of a service is another relevant consideration. For example, a service targeted at children will likely have a higher risk profile than services targeted at adults. In addition to children, online harms can disproportionately impact high-risk users. It should be considered whether a service is more commonly used among other at-risk groups or is used in ways to increase harm for disadvantaged or marginalised groups.
Number of active end-users	The number of active end-users of a service may be relevant to its risk profile. For example, an end-user is more likely to be exposed to harmful content on a service with a large number of active end-users.

	However, the number of active end-users is not always relevant to a service's risk profile. For example, small platforms are often misused for the purposes of distributing content such as CSEM.
Potential for virality	The enabling of rapid and widespread sharing or amplification of material online, which may lead to the viral spread of harmful material, is another relevant consideration when assessing risk. This risk may be more relevant to certain types of online harms than others. For example, while CSEM is unlikely to go viral given the risks inherent in sharing it, pro-terror material is more likely to be shared widely, as we saw after the March 2019 live-streamed attack in Christchurch.

Devices

Type of device being manufactured, installed, repaired or maintained	Device type is a relevant consideration. For example, devices such as laptops, phones, tablets, desktop computers, gaming systems and smart TVs are known to be most commonly used to go online and therefore will have a higher risk profile for online content harms than other internet-enabled devices. Devices that are not designed for general internet browsing or which are unlikely to be used for general internet browsing (for example, connected cars) and/or are located in shared spaces (for example, a smart fridge located in a kitchen), will likely carry a lower risk profile.
Nature of the user	This means the nature of the user of a device. For example, a device targeted at children may have a higher risk profile than devices targeted at adults.

In determining what compliance measures are appropriate for industry participants who manufacture, supply, maintain or install devices, it will also be necessary to consider the role of the industry participant in the supply chain.

For example, compliance measures, such as the provision of information about class 1 and class 2 material or filtering software, should apply at the critical points in the supply chain where people are most receptive to receiving safety information and industry participants are best equipped to provide it. eSafety would expect manufacturers and suppliers (such as retail outlets) who sell directly to the public to provide this information at the point of purchase, as this allows customers to use the information when setting up a new device.

It may be less effective for industry participants who maintain devices (such as those offering mobile phone and computer repairs) to provide this information when undertaking routine repairs (such as screen replacement).

Example of suggested approach

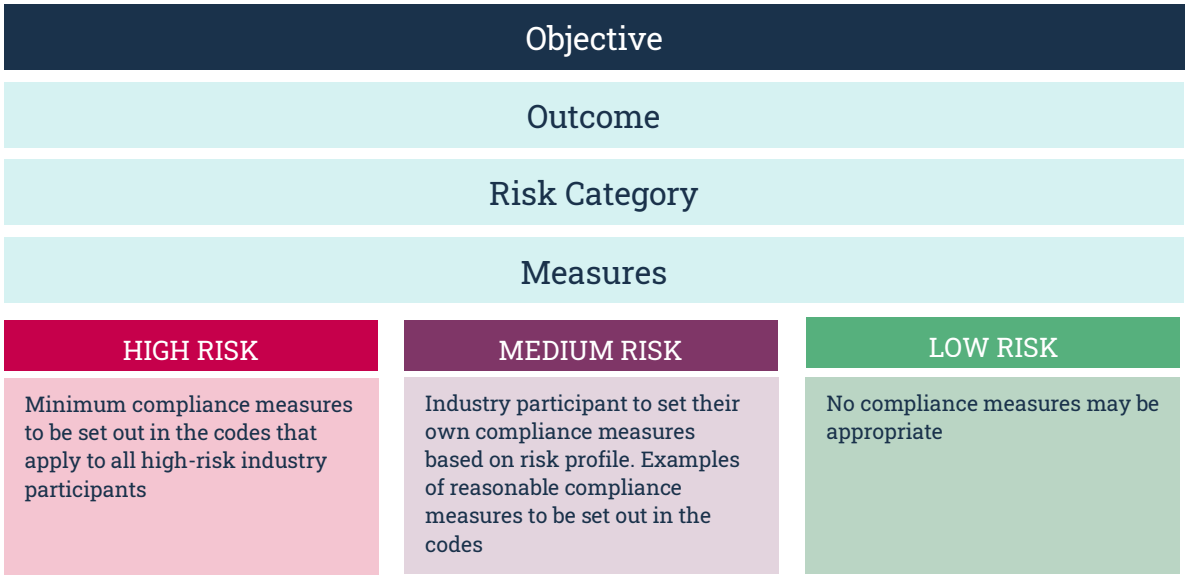
Under the codes, an outcome may require industry participants to proactively detect and prevent access and exposure to, and distribution and online storage of, class 1 – 1A material.

An industry participant whose services and or/devices carry a high risk of facilitating access or exposure to this material should be required, at a minimum, to meet robust, minimum compliance measures, such as a requirement to automatically pre-moderate content. These minimum compliance measures should be set out in the codes.

An industry participant whose services and or/devices carry a medium risk of facilitating access or exposure to this material will determine its own compliance measures. It may, for example, adopt human moderation in response to content reports. The codes should contain examples of measures that could be implemented to help guide participants.

In some other cases, it may be appropriate for the industry participant to take no action in respect of a certain outcome or under the codes overall. eSafety expects this to be the case for a majority of websites, where risk with respect to class 1 and class 2 material is low.

Suggested structure for the codes



Overall, eSafety believes an outcomes- and risk-based approach, supported by minimum compliance measures, is most likely to:

- achieve a balance between flexibility and enforceability, while also providing industry certainty and clarity about required outcomes, priorities and compliance standards
- ensure entities are responsive and accountable for achieving outcomes
- achieve reasonable and proportionate regulatory measures informed by risk.

Codes development

Position 5: Industry associations will prepare all codes for registration by July 2022 or adopt a phased approach to codes development. Under the phased approach, codes dealing with the most harmful content must be lodged for registration by July 2022, and codes dealing with content which is inappropriate for children must be lodged for registration by December 2022.

The Act requires that the eSafety Commissioner make reasonable efforts to ensure that, for each section of the online industry, an industry code is in place within six months of commencement. Industry may choose to prepare and register all codes by July 2022 in line with this provision of the Act.

However, eSafety encourages industry to adopt a two-phased approach to industry codes, where separate codes are developed for:

- high end class 1 material (1A and 1B), including CSEM and pro-terror content
- online pornography and class 2 material.

The two phases have different public policy purposes.

The first phase focuses on all class 1 material except online pornography. These codes will prevent or limit access to material that can be harmful for people of any age. These codes would need to be lodged for registration by July 2022, meeting the Act's expectations for registration of a code for each industry section within six months of commencement.

The second phase will focus on all forms of online pornography and other high impact content that falls within class 2. The purpose of these codes is to prevent children from accessing age-inappropriate content. These codes must be lodged for registration by December 2022.

Class	Material	Phased Codes approach
Class 1- 1A	CSEM	July 2022
	Pro-terror content	
	Extreme crime and violence	
Class 1-1B	Crime and violence	
	Drug-related content	
Class 1 - 1C	Online pornography	December 2022
Class 2 - 2A	Online pornography	
Class 2 - 2B	Online pornography	
	Other high impact content	

This phased approach:

- focuses efforts on preventing or mitigating the most serious material as a priority, so that meaningful measures can be implemented to prevent or limit access and exposure to, and distribution of, this material
- allows industry to leverage the extensive consultation that will be undertaken in the coming months under the age verification roadmap, including engagement with key stakeholders in the adult industry, who will be affected by codes governing online pornography. This may result in greater support and compliance
- facilitates public understanding of the codes, by framing the codes through commonly understood terms
- is consistent with the legislative requirements of the Act.

Position 6: Industry associations will limit the number of codes developed.

While the Act requires a code to be registered for each section of the online industry, the Act does not specify the number of codes required. One code could apply to all eight industry sections, or multiple codes could be registered.

eSafety believes an approach which results in fewer codes is likely to be most effective and efficient. If the codes are developed in two phases, as outlined in position 5, this could include the development of one code per phase which contains chapters for different industry sections. Alternatively, one code could be developed for class 1 material and one code could be developed for class 2 material.

Matters relating to some industry sections may also be able to be combined and covered in the same code or same chapter of a code. For example, social media services and relevant electronic services offer similar functionalities relevant to the access and distribution of class 1 and class 2 material (such as the ability to chat with users or share images or videos). The issues to be addressed in the codes are therefore likely to be the same or similar across these industry sections.

Within a code, or chapter of a code, which covers multiple industry sections, minimum compliance measures could apply by functionality, rather than to a particular industry section. For example, minimum compliance measures could apply to all services that offer the ability for users to chat with other users, regardless of whether that service is a social media service or relevant electronic service.

Minimising the number of codes helps avoid duplication, gaps in safeguards and/or inconsistent approaches where industry participants fall into multiple industry sections. It will also likely lead to less confusion amongst the Australian public.

If more than one code is developed, an overarching framework must be put in place to ensure consistency across codes.

Registration

Position 7: Industry associations will engage widely with participants within their industry section(s) to ensure they adequately represent each section covered by a code.

In order to register a code, the eSafety Commissioner must be satisfied that the industry association (or industry associations) that drafted the code represents the particular section (or sections) of the online industry to which the code applies.

eSafety acknowledges that the eight industry sections cover a diverse set of online activities, and that some industry sections capture a wide range of services and devices. The members of an industry association may fall within multiple industry sections. Some industry participants may be members of multiple industry associations.

Given these overlaps, there are various ways in which industry associations may choose to draft codes:

- a code could be developed by a single industry association, covering a single or multiple online industry sections
- a code could be developed by multiple industry associations working together (including by forming a new drafting body), covering a single or multiple online industry sections.

eSafety expects that the arrangements for code drafting will be determined by the industry associations, noting its preference for a limited number of industry codes.

eSafety will be taking a practical and balanced approach to assessing representation that includes the following considerations:

- **Breadth and depth of representation**

  Representation may be a matter of both breadth (representing different types of participants) and depth (representing a reasonable number of participants). This does not mean that every participant, or every type of participant, must necessarily be accounted for, but that there is sufficient representation of participants, such that the industry association could be said, broadly, to be speaking on behalf of the section as a whole.

  eSafety considers that this requires an industry association to represent industry participants of different sizes, including small or early-stage companies (0-49 employees worldwide), mid-tier companies (50-249 employees worldwide) and enterprise companies (250+ employees worldwide).

  Representation must also include those industry participants whose services and devices present the greatest risk of access and exposure to, and distribution of, class 1 and class 2 material, and who are therefore most likely to be impacted by the codes. This will include, for example, services whose primary purpose is to publish class 2 content (such as online pornography websites).

  eSafety is of the view that the greatest breadth and depth of representation is most likely to be achieved through multiple industry associations working together to draft a code. In line with its preference for fewer codes, eSafety encourages groups of industry associations to

come together and register a code that covers the widest range of sections possible.

- **Representation of a specific sub-section**

  A body representing only a distinct part of an industry section will not be able to demonstrate that they represent the whole section. eSafety does not intend to register codes for subsections of industry sections.

- **A continuum of representation**

  eSafety considers that there is a continuum of representation, which does not necessarily require membership of an industry association.

  At one end of the spectrum, there is consultation with industry participants. The Act contains a separate but related requirement for industry consultation at section 140(1)(f). While consultation does not necessarily equate to representation, the more robust a consultation process is with participants of an industry section, the more it can contribute to an industry association's claim to represent that section.

  At the other end of the spectrum, there is membership. eSafety can be confident that an industry association represents the industry participants which are its members.

  In between, there are a range of engagement processes or arrangements that may demonstrate representation, and eSafety encourages industry to think about how best to achieve this. For example, industry associations may reach out to industry participants for input prior to drafting a code, or industry participants may feed into the drafting process at various stages. Industry participants may agree to the industry association putting forward the draft code on their behalf.

  If an industry association is able to demonstrate that a range of industry participants contributed to the drafting of the code and indicated their agreement with the code, eSafety can be reasonably assured that the association speaks on their behalf.

eSafety has worked closely with a number of industry associations to date and intends to continue working with these associations (and others) to advance the codes development process. This includes ensuring these associations engage with a wide variety of participants across sections, including actively seeking out small and start-up participants and those industry participants whose services and devices carry risks relating to class 1 and class 2 material.

Furthermore, eSafety also encourages industry participants who are currently not part of an industry association to consider working through a representative industry association to provide input to these codes. This will provide an opportunity for those who have not previously engaged in the Australian co-regulatory environment to be involved in the process. It will also enable these industry participants to be informed by the regulatory experience of those associations and participants that may have taken part in similar codes development processes.

Industry standards

In certain circumstances, including where there is no industry association to represent an industry section or where an existing industry association fails to develop a code, eSafety may be required to impose a standard. The circumstances where eSafety may impose a standard are outlined in **3.0 Regulatory responses to harmful online content.**

Position 8: Industry associations will conduct meaningful industry and public consultation.

Prior to registration, an industry association developing an industry code must undertake two consultation processes. It must publish a draft of the code and invite submissions:

- from members of the public
- from industry participants in the applicable section of the online industry.

Each consultation must run for at least 30 days, and it is possible for the two consultation periods to occur concurrently if the industry association considers it appropriate. If a subsequent period of consultation is required, this time period must be reasonable, but does not need to be at least 30 days.

General principles

eSafety expects industry associations to conduct meaningful consultation processes: consultation should be genuine, widely accessible and transparent. This will help ensure that the industry association is aware of the potential impact of the code on interested parties. It will also help ensure the code adequately considers and addresses any relevant community concerns or needs for safeguards.

Appropriate forms of consultation may include working groups, focus groups, surveys or web forums. A combination of methods of consultation may be the best strategy to ensure effective consultation with interested parties.

To ensure the consultation process is widely accessible, input should be proactively sought from a variety of stakeholders and notice of the consultation processes should be widely published to encourage submissions.

In order for the consultation to be transparent, eSafety considers that all submissions received as part of the codes development process should be genuinely considered and made available on the website of the industry association.

Public consultation

In relation to public consultation, relevant parties in this process may include, but are not limited to:

- consumer groups
- civil society groups
- community legal and advocacy groups
- representatives from academia
- children and young people
- parents, carers, teachers and educators (including their representative groups)
- users of the services and devices (including content creators impacted by the codes)
- digital rights groups
- women's advocacy groups
- domestic and family violence groups
- groups representing sex workers
- safety tech sector.

Some stakeholders may form part of both public and industry consultation. eSafety also encourages industry associations to consider the voices and experiences of diverse and at-risk groups in developing industry codes.

Any invitation to provide input on a draft code should provide easy and accessible access to the draft, use plain language to explain the purpose of

the code and clarify the key issues involved. The invitation should also explain the ways in which the public can contribute to the codes development process.

Industry consultation

Industry associations must take steps to ensure eSafety can be satisfied they have consulted adequately. As outlined in Position 7, in assessing representation as part of the code registration process, eSafety will take into consideration the scale and breadth of the industry consultation conducted.

Position 9: Industry associations will engage with eSafety throughout the codes development process.

The Act requires the Commissioner to be consulted about the development of industry codes.

eSafety has already worked closely with a number of industry associations to scope relevant issues and produce this position paper. eSafety appreciates the collaborative and constructive way industry has engaged with the codes development process so far and encourages other interested industry associations to take part.

eSafety expects industry associations to continue to engage with eSafety throughout the codes process. This engagement will assist in tracking the progress of the codes and ensure any areas of overlap or duplication, or other concerns, are addressed early. It will also promote consistency and coordination across codes and ensure the registration process runs smoothly.

Industry have indicated that they may establish a Steering Group, made up of multiple industry associations, to progress the drafting of codes. eSafety strongly supports the establishment of a Steering Group which could draft a code that covers the widest range of industry sections possible.

Administering and reviewing the codes

Successful administration of the codes will require:

- robust enforcement of minimum compliance measures
- a consistent program of evaluation, review and reporting to establish that outcomes are being achieved.

This includes effective mechanisms and processes to handle complaints about code breaches.

Compliance and enforcement

Position 10: Industry participants will handle reports and complaints about class 1 and class 2 material and codes compliance in the first instance. eSafety will act as a 'safety net' if resolution of a complaint is not satisfactory.

The role of eSafety

eSafety's approach to codes investigations, compliance and enforcement will be set out in regulatory guidance.

At this stage, eSafety's view is that industry participants should have the opportunity to first address the following matters internally through a transparent and responsive process:

- reports of class 1 and class 2 material
- complaints about the handling of a class 1 or class 2 report
- complaints about codes compliance.

It is intended that eSafety will generally act as a 'safety net' if resolution of a complaint is not satisfactory to the complainant.

Further, eSafety may choose to exercise its discretion not to investigate certain online content complaints in the first instance where there is a code in place that can address those complaints.

Where a complaint is unable to be resolved, or where a complainant is dissatisfied with the way their complaint is handled, either:

- the person can lodge a complaint with eSafety, or
- the industry participant may refer the complaint to eSafety (where the codes contain such an option).

Conversely, there may be cases where eSafety proactively elects to use its investigative powers to determine whether an industry participant has breached an industry code. This could include, for example, where eSafety has received a number of online content complaints from the public about the availability of class 1 or class 2 material which may indicate that an industry participant has a systemic issue with codes compliance.

The role of industry associations

eSafety encourages the industry associations who draft the codes to consider the role they will play in compliance and enforcement and review and evaluation of the codes. Industry associations may also wish to formalise a representative committee or other third-party body to oversee administration of a code.

Industry associations (or any third-party body) could, for example:

- triage, review or manage complaints made against the code – but this approach may not be possible or appropriate for all complaints, particularly given complaints may include links to, or descriptions of, class 1 material
- monitor compliance with codes by, for example, undertaking periodic audits of participants (such as annually or quarterly)
- educate industry participants about the role of the codes
- act as an avenue for the sharing of new technologies, data, factchecks and initiatives between industry participants
- review the codes to ensure they are up to date and working well.

Information or reports from these processes could be provided to eSafety. eSafety anticipates working closely with industry to ensure appropriate administration of the codes.

Reporting

This paper proposes that strengthening transparency of, and accountability for, class 1 and class 2 material should be an objective of the codes.

While eSafety does not expect the codes to require regular reporting to eSafety, it does expect high-risk industry participants in particular to publish annual reports about codes compliance.

eSafety also encourages industry to consider what other reporting processes may be appropriate to ensure codes compliance is reviewed and evaluated on an ongoing basis.

Position 11: The codes will include a review mechanism.

eSafety expects that the codes will include a statement about how and when they will be reviewed. This should include a mandated review after an initial time period. For example, the codes could be reviewed at 12 months, and thereafter every three years.

Any review should consider:

- continued relevance of the code requirements
- developments that have created gaps in the codes that should be filled
- areas that have caused confusion for industry participants
- how industry members have complied with the codes, including results of any compliance monitoring and insights from complaints handling
- how successful or unsuccessful the codes have been in preventing and mitigating harm
- the public's understanding and response to the codes.

5.0 Preferred codes model

When registering a code, eSafety must be satisfied that the code provides appropriate community safeguards for the matters of substantial relevance to the community which are covered by the code.

To facilitate the development of robust codes, eSafety has developed a proposed outcomes-based model. When registering a code, eSafety will consider the extent to which the code aligns with this model.

This model includes:

1. the purpose of the codes
2. the objectives of the codes
3. the outcomes of the codes
4. examples of the kinds of measures that industry participants could put in place to meet the outcomes of the codes. eSafety encourages industry to:
    a. include minimum compliance measures in the codes for industry participants whose services and/or devices are considered high risk; and
    b. include examples in the codes as to what measures industry participants could otherwise adopt in order to meet the objectives and outcomes of the codes. These industry participants must determine what measures are appropriate and proportionate to meet the objectives and outcomes of the codes, considering the risk profile of the industry participant's services and devices.

Generally, the proposed outcomes apply to both class 1 and class 2 material. However, some outcomes treat certain types of class 1 and class 2 material differently. This recognises that:

- class 1 and class 2 material include a range of online material associated with different types and degrees of harm
- the codes should, where possible, give Australian adults autonomy and control over the online material they can access, create and share.

For example, class 1 – 1A material is material that is seriously harmful and generally should not be accessible online. Class 1 – 1B material is also harmful but may be appropriate for adults to access provided suitable limitations are in place (such as a warning notice). eSafety encourages industry to approach

the different types of material covered by the codes in a manner which is in the spirit of what the objectives and outcomes are trying to achieve.

eSafety also acknowledges that industry participants may elect to treat certain types of class 1 and class 2 material differently. For example, some industry participants may choose to prevent all persons from accessing or distributing pornography on a service, while others may only prevent children from accessing this material. The approach taken by an industry participant with respect to class 1 and class 2 material will affect how it complies with the outcomes. For example, a service which does not allow access to class 1 or class 2 material will not necessarily need separate reporting and complaints mechanisms for each type of material.

Each code should also include:

1. an opening statement that enables readers to understand the purpose of the code, what it seeks to achieve, the type of material that is covered by the code and which industry sections are subject to the code
2. a glossary which explains key concepts. However, terms that are already defined under the Online Safety Act should not be redefined. Each code should include the same glossary for consistency.

Preferred outcomes-based codes model

<table>
<tr><th colspan="4">PURPOSE: To ensure that participants of the online industry provide appropriate community safeguards for Australians in relation to class 1 and class 2 material</th></tr>
<tr><td colspan="4">OBJECTIVE 1: Industry participants will take proactive steps to create and maintain a safe online environment</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Material prevention or restriction</td><td>Industry participants proactively detect and prevent:*
access or exposure to,
distribution of, and
online storage of,
Class 1 - 1A[28] material</td><td>Industry participants proactively prevent or limit:
access or exposure to, and
distribution of,
Class 1 - 1B[29] material **</td><td>Industry participants proactively:
prevent access or exposure to, and distribution of, or
prevent children from accessing or being exposed to,
Class 1 - 1C[30] and Class 2[31] material ***</td></tr>
<tr><td>Hosting</td><td colspan="3">Industry participants do not host class 1 and class 2 – 2A[32] material in Australia. Industry participants who host Class 2 – 2B[33] material in Australia prevent children from accessing, or being exposed to, that material</td></tr>
<tr><td>Industry cooperation</td><td colspan="3">Industry participants consult, cooperate and collaborate with other industry participants in respect of the removal, disruption and/or restriction of class 1 and class 2 material</td></tr>
<tr><td>Cooperation with Commissioner</td><td colspan="3">Industry participants communicate and cooperate with the eSafety Commissioner in respect of matters relating to class 1 and class 2 material, including complaints</td></tr>
<tr><td colspan="4">OBJECTIVE 2: Industry participants will empower people to manage access and exposure to class 1 and class 2 material</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Tools and information</td><td colspan="3">Industry participants provide tools and/or information to limit access and exposure to class 1 and class 2 material</td></tr>
<tr><td>Reporting of material</td><td colspan="3">Industry participants provide robust and effective reporting and complaints mechanisms for class 1 and class 2 material</td></tr>
<tr><td>Report handling</td><td colspan="3">Industry participants effectively respond to reports and complaints about class 1 and class 2 material</td></tr>
<tr><td colspan="4">OBJECTIVE 3: Industry participants will strengthen transparency of, and accountability for, class 1 and class 2 material</td></tr>
<tr><td colspan="4">Outcomes:</td></tr>
<tr><td>Public policies</td><td colspan="3">Industry participants provide clear and accessible information about class 1 and class 2 material</td></tr>
</table>

[28] This includes child exploitation material, pro-terror content and extreme crime and violence.
[29] This includes crime and violence and drug-related content.
[30] This includes online pornography (RC).
[31] This includes online pornography (X18+), online pornography (R18+) and other high impact content.
[32] This includes online pornography (X18+).
[33] This includes online pornography (R18+) and other high impact content.

Public reporting	Industry participants publish periodic reports about class 1 and class 2 material and codes compliance

*Industry participants should take reasonable steps to proactively prevent access or exposure to, and distribution and online storage of, this material. However, failure to prevent access or exposure to, and distribution and online storage of, this material, does not necessarily indicate that there has been a codes breach. Where this material is accessible, 'prevention' includes quick removal of this material.

** At a minimum, industry participants must proactively limit access or exposure to, and distribution of, class 1 – 1B material.

***At a minimum, industry participants must proactively prevent children from accessing, or being exposed to, class 1 - 1C and class 2 material.

Objectives

The objectives of the codes have been devised from section 138(3) of the Act, which provides a list of examples of matters that may be dealt with by industry codes. Generally, eSafety expects industry to have regard to the matters set out at 138, to the extent that the matter applies to the relevant industry section.

OBJECTIVE 1	**Industry participants will take proactive steps to create and maintain a safe online environment** To create a safe online environment, each code should not only address access and exposure to class 1 and class 2 material, but also the risks and harms associated with distribution of this material, including live-streaming. eSafety also expects industry participants to consider their relationships and opportunities for collaboration with other industry participants, locally and globally, and with the Commissioner.	138(3)(a) – (c) 138(3)(d) 138(3)(f) 138(3)(k) 138(3)(o)-(q)	138(3)(zc)-(ze) 138(3)(zf) 138(3)(zh)-(zj)
OBJECTIVE 2	**Industry participants will empower people to manage access and exposure to class 1 and class 2 material** Each code should seek to empower and equip people to limit their access and exposure to class 1 and class 2 material through tools and information. People should also be able to easily report class 1 and class 2 material, and industry participants must effectively respond to these reports.	138(3)(g) 138(3)(h) 138(3)(i) 138(3)(j) 138(3)(r)-(t)	138(u) 138(3)(v)-(x) 138(3)(y)-(za) 138(3)(zb)
OBJECTIVE 3	**Industry participants will strengthen the transparency of, and accountability for, class 1 and class 2 material**	138(3)(e) 138(3)(l)-(n) 138(3)(zg)	

	Each code should require transparency and accountability from industry participants.	

Outcomes of the codes

Proactive steps

OBJECTIVE 1: Industry participants will take proactive steps to create and maintain a safe online environment

Outcome 1: Industry participants proactively detect and prevent:

- access or exposure to,
- distribution of, and
- online storage of,

class 1 - 1A material.

Industry participants will have scalable and effective policies, procedures, systems and technologies in place to proactively detect and prevent:

- access or exposure to,
- distribution of, and
- online storage of

class 1 - 1A material.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Ongoing investment in, and development and use of, tools to detect, moderate and report material (for example, through the use of hashing, machine learning, artificial intelligence or other safety technologies)
- Development and use of effective moderation practices and procedures (for example, automatic pre-moderation, proactive machine monitoring, human monitoring, hybrid moderation, appointed community moderators and community moderation) to take action against harmful content and activity, including through warning account-holders, suspending or removing accounts, removing content and deindexing of search results
- Default settings for services marketed to children which are set to the highest possible privacy and safety level at registration or sign up (for example, access to device hardware such as cameras and microphones is limited and photos, location, friends lists, profile information and chat functions are only accessible to approved contacts. This might also include safety settings such as safe search mode on by default and measures which would prevent comingling)
- Standard operating procedures which include clearly specified channels for escalating and/or reporting unlawful and harmful material, including to law enforcement, child protection or relevant authorities

- Development and use of filtering, labelling and classification processes and technologies to prevent, limit and mitigate access or exposure to class 1 - 1A material.

Outcome 2: Industry participants proactively prevent or limit:

- access or exposure to, and
- distribution of,

class 1 - 1B material.

Industry participants will have scalable and effective policies, procedures, systems and technologies in place to proactively prevent or limit:

- access or exposure to, and
- distribution of,

class 1 - 1B material.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile), in addition to the measures set out under Outcome 1:

- Age gating through age verification or age assurance mechanisms
- Interstitial notices
- Warning labels, warning/notice screens
- Downlisting or deprioritising content
- Quarantining
- Image/text/audio masking
- Reducing promotion and reach within algorithmic systems, including recommendation algorithms and choice architecture
- Internal policies and procedures to proactively monitor, assess, investigate, and audit content within algorithmic systems.

Outcome 3: Industry participants proactively:

- prevent access or exposure to, and distribution of, or
- prevent children from accessing, or being exposed to

class 1 - 1C and class 2 material.

Industry participants will have scalable and effective policies, procedures, systems and technologies in place to proactively:

- prevent access or exposure to, and distribution of, or
- prevent children from accessing, or being exposed to,

class 1 - 1C and class 2 material.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Implementation of age verification or age assurance mechanisms
- Safety settings such as safe search mode are turned on by default
- Internal policies and procedures that include safety risk and impact assessments, and safety review processes that specifically consider users aged under 18.

Outcome 4: Industry participants do not host class 1 and class 2 – 2A material in Australia. Industry participants who host class 2 – 2B material in Australia prevent children from accessing, or being exposed to, that material.

Industry participants will have scalable and effective policies, procedures, systems and technologies in place to ensure that class 1 and class 2 – 2A material is not hosted in Australia. Industry participants who host 2 – 2B material in Australia will have scalable and effective policies, procedures, systems and technologies in place to prevent children from accessing, or being exposed to, that material.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Robust contracts with customers which stipulate hosting requirements
- Reporting functionality
- Hash scanning technologies that work within customer storage environments or cloud infrastructure to detect CSEM.

Outcome 5: Industry participants consult, cooperate and collaborate with other industry participants in respect of the removal, disruption and/or restriction of class 1 and class 2 material.

Industry participants will have effective and scalable policies and procedures in place to facilitate consultation, cooperation and collaboration with other industry participants in respect of the removal, disruption and/or restriction of class 1 and class 2 material, as well as accounts associated with this material.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Procedures for the sharing of data, information, intelligence and other relevant insights (for example, hash information and URLs relating to class 1 - 1A material)
- Procedures to expeditiously notify other industry participants about viral class 1 or class 2 material (for example, a viral suicide video) and coordinate cross platform/cross sector action and removal of material
- Proactive engagement with local and global industry and multi-stakeholder communities, coalitions and alliances to share information and best practices (for example, open sourcing detection and moderation technologies, supporting research and innovation and contributing to cross-sector online safety groups and initiatives).

Outcome 6: Industry participants communicate and cooperate with the eSafety Commissioner in respect of matters relating to class 1 and class 2 material, including complaints.

Industry participants will have effective and scalable policies and procedures in place which ensure communication and cooperation with the eSafety Commissioner with respect to matters about class 1 and class 2 material, including complaints.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Policies and procedures directed to the timely referral to the Commissioner of complaints about matters relating to class 1 and class 2 material or codes compliance, where the complainant is dissatisfied with the way in which their complaint was dealt with under a code
- Policies and procedures which ensure the Commissioner receives timely updates regarding technological developments which could have a positive or negative effect on the safety of children.

User empowerment

OBJECTIVE 2: Industry participants will empower people to manage access and exposure to class 1 and class 2 material

Outcome 7: Industry participants provide tools and/or information to limit access and exposure to class 1 and class 2 material.

Industry participants will provide people with range of technical tools and/or information to limit their access and exposure, and the access and exposure of children in their care, to class 1 and class 2 material.

Examples of measures through which this outcome could be implemented (depending on service or device provided and associated risk profile)**:**

- Provision of a range of privacy and safety settings, accompanied by clear and accessible guidelines about the use and effect of such settings
- Provision of parental/carers companion apps and/or controls
- Provision of tools which enable users to rate material
- Provision of tools to block users, accounts or otherwise control interactions on a service
- Development and use of filtering, labelling and classification processes and technologies to prevent, limit and mitigate access and exposure to class 1 and class 2 material, accompanied by clear and accessible guidelines and information about the use and effect of filters and other tools
- Provision of advice on how to limit access and exposure to class 1 and class 2 material
- Provision of information to parents and carers about how to supervise and manage children's access and exposure to class 1 and class 2 material.

Outcome 8: Industry participants provide clear and effective reporting and complaints mechanisms for class 1 and class 2 material.

Industry participants will provide clear, easily accessible and effective reporting tools in respect of class 1 and class 2 material, as well as associated user accounts. Industry participants will provide clear, easily accessible and effective complaints mechanisms to address complaints about the handling of reports about class 1 and class 2 material and codes compliance.

Examples of measures through which this outcome could be implemented (depending on service or device provided and associated risk profile):

- Provision of on-platform flagging and reporting tools which are able to be accessed within a service (such as in-app or in-chat) and not on a separate platform or website
- Provision of off-platform flagging and reporting tools (for example, email address)
- Provision of reporting tools which:
  - are easy to access
  - are accompanied by plain language instructions on how to use the reporting tools, as well as an overview of the reporting process
  - ensure persons can report content without re-exposure
  - ensure persons can report content anonymously or opt-out of further engagement (for example, if the content is triggering or to prevent further harm)
  - provide users with direct links to third party support services.

- Provision of easy to access mechanisms to make complaints about the handling of reports about class 1 and class 2 material and codes compliance, accompanied by plain language instructions.

Outcome 9: Industry participants effectively respond to reports and complaints about class 1 and class 2 material.

Industry participants will have scalable and effective policies, procedures, systems and technologies in place to effectively respond to reports about class 1 and class 2 material, as well as associated user accounts. Industry participants will have scalable and robust policies, procedures, systems and technologies in place to effectively respond to complaints about the handling of reports about class 1 and class 2 material and codes compliance.

Examples of measures through which this outcome could be implemented
(depending on service or device provided and associated risk profile):

- Automated, human or hybrid triaging of reports and complaints
- Policies and procedures which:
  - ensure reports are assessed and material removed or otherwise actioned (if necessary) within appropriate timeframes, based on the level of harm posed by the material
  - ensure ongoing engagement with the person who made the report to advise of progress and provide opportunities for review
  - ensure notification of outcomes to the person who made the report
- Policies and procedures which:
  - give persons the opportunity to request and receive a review of how a report was handled
  - ensure complaints about the handling of a class 1 or class 2 material report are assessed and actioned within appropriate timeframes
  - ensure complaints about codes compliance are assessed and actioned within appropriate timeframes
  - ensure ongoing engagement with the person who made the complaint to advise of progress and provide opportunities for review.
- Policies and procedures directed to the timely referral to the Commissioner of complaints about matters relating to class 1 and class 2 material or codes compliance, where the complainant is dissatisfied with the way in which their complaint was dealt with under a code
- Standard operating procedures which include clearly specified channels for escalating and/or reporting unlawful and harmful material, including to law enforcement, child protection or relevant authorities.

Transparency and accountability

OBJECTIVE 3: Industry participants will strengthen transparency of, and accountability for, class 1 and class 2 material

Outcome 10: Industry participants provide clear and accessible information about class 1 and class 2 material.

Industry participants will publish easily accessible and plain language policies, procedures and guidelines that set out how they handle class 1 and class 2 material. Industry participants will also provide information about the safety issues associated with class 1 and class 2 material.

Examples of measures through which this outcome could be implemented (depending on service or device provided and associated risk profile):

- Publication of terms of service, acceptable use policies, community standards and/or other policies, procedures or guidelines which:
  - inform users about the types of materials that are prohibited from being accessed and/or distributed via the relevant industry participant's service, and the procedures, systems and technologies in place to deal with these materials, including class 1 - 1A material
  - inform users about online safety issues in respect of class 1 and 2 material, including information for parents/carers about how to supervise and control children's access and exposure to class 1 and class 2 material, and
  - provide information about the role and functions of the eSafety Commissioner, including how to make a complaint to eSafety
  - are easy to access at all points of the user experience, including registration, account creation, first use and at regular intervals
- Establishment of a dedicated section to house online safety information, such as a safety centre
- Provision of information at the point of purchase of a device, including information about online safety issues in respect of class 1 and 2 material and information for parents/carers about how to supervise and control children's access and exposure to class 1 and class 2 material. Information about the role and functions of eSafety, including how to make a complaint to eSafety, should also be provided.

Outcome 11: Industry participants publish annual reports about codes compliance.

Industry participants will publish annual reports about class 1 and class 2 material and their compliance with the codes. These reports could include:

- Meaningful analysis in respect of:
  - the number of reports received about class 1 and class 2 material
  - the number of complaints received in respect of handling of reports in respect of class 1 and class 2 material
  - the number of complaints received in respect of codes compliance.
- The measures taken by the industry participant to comply with the outcomes of the codes
- The effectiveness of the measures taken by the industry participant to comply with the outcomes of the codes, including, for example, the effectiveness of detection, moderation and remediation efforts

- Data and information on safety innovations, investments and third-party engagements (such as cooperation efforts).

Facilitation of class 1 – 1A material

As set out in Position 1, the risks and harms associated with class 1 and class 2 material are not limited to access and exposure, and extend to practices of manipulation, coercion, enticement and exploitation that act as precursors to the production and distribution of, for example, CSEM.

As far as practicable, eSafety expects that the codes will address the facilitation of class 1 – 1A material.

The codes should include an outcome requiring industry participants to have scalable and effective policies, procedures, systems and technologies in place to proactively detect and prevent or mitigate contact between users which could facilitate the production of class 1 - 1A material. For example, contact or messaging involving the grooming of a child to facilitate the production of CSEM.

Examples of measures through which this outcome could be implemented (depending on service or device provided and associated risk profile), in addition to the measures set out under Outcome 1, include:

- Default settings on services that are implemented according to the age of the user. Services marketed to children have default settings which are set to the highest possible privacy and safety level at registration or sign up (for example, access to device hardware such as cameras and microphones is limited and photos, location, friends lists, profile information and chat functions are only accessible to approved contacts. This might also include safety settings such as safe search mode on by default and measures which would prevent comingling)
- Use of tools and processes to ensure that users who have been removed from a service or whose accounts are blocked aren't able to re-register for the service (for example, device level blocking, IP blocking, client-side cookies, geofencing, rate limiting, formal identity validation and verification).

6.0 Registration process

Once a code is finalised, an electronic copy must be lodged with eSafety, accompanied by supporting documentation.

Both the code and the supporting documentation will be used by eSafety to assess whether the code should be registered. This means the supporting documentation will need to provide a range of information, including:

- the name of the industry association(s) lodging the code
- an explanation of the industry section(s) the code is intended to cover
- evidence demonstrating how the industry association(s) represents the industry section(s). This could include:
    - a list of members
    - a list of other represented industry participants and how they have been engaged in the development of the code
- an explanation as to how the code provides appropriate community safeguards for matters of substantial relevance to the community, including how the code aligns with eSafety's 11 positions set out in **4.0 eSafety positions on codes development** and the model set out in **5.0 Preferred codes model.** The association should also explain how any deviations achieve the policy intent of these positions
- details of the industry and public consultation that was undertaken during the drafting of the code. Information could include a summary of how the consultations were publicised and run, a summary of the feedback received and any changes made to the draft code to reflect it.

eSafety expects that it will require at least four weeks to review any code(s) lodged for registration. All approved codes will be maintained in an electronic Register that will be publicly available on the eSafety website.

7.0 Next steps

eSafety will continue to engage with industry to guide codes development.

eSafety acknowledges the tight timeframes for codes drafting and registration.

To ensure the codes develop in a timely manner, eSafety will work with industry to develop:

- **timelines for codes development and registration (including milestone dates) within 30 days of the release of this paper.**
- **a framework for the development of consistent codes within 45 days of the release of this paper.**

If industry does not adopt eSafety's preferred two-phased approach, it will need to lodge robust class 1 and class 2 codes, covering all eight sections on the online industry, within six months of the commencement of the Act. If industry adopts a two-phased approach, the first codes need to be lodged for registration by July 2022, and the second codes by December 2022.

eSafety will continue to provide further information and guidance as the code process develops.

Appendix A: Timeline

A timeline of online content regulation

- Schedule 5 (Online services) inserted into Broadcasting Services Act 1992 (Cth) (BSA) through the Broadcasting Services Amendment (Online Services) Act 1999 (July 1999)
- Codes registered for industry co-regulation in areas of internet and mobile content (consisting of three codes), pursuant to Schedule 5 of the BSA (May 2005)
- Schedule 7 (Content services) inserted into BSA through the Communications Legislation Amendment (Content Services) Act 2007 (July 2007)
- Code registered for industry co-regulation in the area of content services, pursuant to Schedule 7 of the BSA (July 2008)
- Enhancing Online Safety Act 2015 (Cth) commenced (July 2015)
- Enhancing Online Safety Act 2015 (Cth) and Online Content Scheme reviewed (Briggs Review) (report released February 2019)
- Consultation started on a new Online Safety Act (December 2019)
- Online Safety Bill released for public consultation (December 2020)
- Senate inquiry into Online Safety Bill held (March 2021)
- eSafety commenced consultation with online industry on new codes (May 2021)
- Online Safety Act received Royal Assent (July 2021)
- Industry codes position paper released (September 2021)
- Online Safety Act 2021 (Cth) (including new Restricted Access System (RAS)) commences (January 2022)
- Registration of new industry codes (July 2022). If the suggested two-phased approach is adopted, the first codes will be registered by July 2022 and the second phase of codes will be registered by December 2022
- Registration of industry standards (where industry codes are not registered for some or all sections of the online industry) (January 2023)

Appendix B: Glossary

This glossary outlines the legislative definition for a number of key terms under the Online Safety Act.

App	Section 5: Definitions Includes a computer program.
App distribution service	Section 5: Definitions Means a service that enables end users to download apps, where the download of the apps is by means of a carriage service.
Designated Internet service	Section 14: Designated internet service (1) For the purposes of this Act, designated internet service means: (a) a service that allows end-users to access material using an internet carriage service; or (b) a service that delivers material to persons having equipment appropriate for receiving that material, where the delivery of the service is by means of an internet carriage service; but does not include: (c) a social media service; or (d) a relevant electronic service; or (e) an on demand program service; or (f) a service specified under subsection (2); or (g) an exempt service (as defined by subsection (3)). (2) The Minister may, by legislative instrument, specify one or more services for the purposes of paragraph (1)(f). Exempt services (3) For the purposes of this section, a service is an exempt service if none of the material on the service is accessible to, or delivered to, one or more end-users in Australia.
Film	Section 5: Definitions Film has the same meaning as in the Classification (Publications, Films and Computer Games) Act 1995.
Hosting service	Section 17: Hosting service For the purposes of this Act, if: (a) a person (the first person) hosts stored material that has been provided on: (i) a social media service; or (ii) a relevant electronic service; or (iii) a designated internet service; and (b) the first person or another person provides: (i) a social media service; or (ii) a relevant electronic service; or (iii) a designated internet service; on which the hosted material is provided;

	the hosting of the stored material by the first person is taken to be the provision by the first person of a hosting service.
Internet carriage service	Section 5: Definitions Means a listed carriage service that enables end users to access the internet.
Internet service provider	Section 19: Internet service providers Basic definition (1) For the purposes of this Act, if a person supplies, or proposes to supply, an internet carriage service to the public, the person is an internet service provider. Declared internet service providers (2) The Minister may, by legislative instrument, declare that a specified person who supplies, or proposes to supply, a specified internet carriage service is an internet service provider for the purposes of this Act. Note: For specification by class, see subsection 13(3) of the Legislation Act 2003.
Relevant electronic service	Section 13A: Relevant electronic service (1) For the purposes of this Act, relevant electronic service means any of the following electronic services: (a) a service that enables end-users to communicate, by means of email, with other end-users; (b) an instant messaging service that enables end-users to communicate with other end-users; (c) an SMS service that enables end-users to communicate with other end-users; (d) an MMS service that enables end-users to communicate with other end-users; (e) a chat service that enables end-users to communicate with other end-users; (f) a service that enables end-users to play online games with other end-users; (g) an electronic service specified in the legislative rules; but does not include an exempt service (as defined by subsection (2)). Note 1: SMS is short for short message service. Note 2: MMS is short for multimedia message service. Exempt services (2) For the purposes of this section, a service is an exempt service if none of the material on the service is accessible to, or delivered to, one or more end-users in Australia.
Social media service	Section 13: Social media service (1) For the purposes of this Act, social media service means: (a) an electronic service that satisfies the following conditions: (i) the sole or primary purpose of the service is to enable online social interaction between 2 or more end-users;

	(ii) the service allows end-users to link to, or interact with, some or all of the other end-users; (iii) the service allows end-users to post material on the service; (iv) such other conditions (if any) as are set out in the legislative rules; or (b) an electronic service specified in the legislative rules; but does not include an exempt service (as defined by subsection (4)). Note: Online social interaction does not include (for example) online business interaction. (2) For the purposes of subparagraph (1)(a)(i), online social interaction includes online interaction that enables end-users to share material for social purposes. Note: Social purposes does not include (for example) business purposes. (3) In determining whether the condition set out in subparagraph (1)(a)(i) is satisfied, disregard any of the following purposes: (a) the provision of advertising material on the service; (b) the generation of revenue from the provision of advertising material on the service. Exempt services (4) For the purposes of this section, a service is an exempt service if: (a) none of the material on the service is accessible to, or delivered to, one or more end-users in Australia; or (b) the service is specified in the legislative rules.

Appendix C: International approaches to codes

European Commission Code of Conduct on Countering Illegal Hate Speech Online (Hate Speech Code)

The 2016 Hate Speech Code is a non-binding, voluntary code designed to prevent and counter the spread of illegal hate speech, including racist and xenophobic content and terrorist propaganda. Facebook, Microsoft, Twitter and YouTube signed the Code in 2016, followed by Instagram, Snapchat, Google+ and Dailymotion in 2018, Jeuxvideo.com in 2019, TikTok in 2020 and LinkedIn in 2021.

The scope of the Hate Speech Code is limited to illegal hate speech.

The Hate Speech Code sets out commitments that signatories agree to follow, including to:

- have rules and community standards that prohibit hate speech and put in place systems and teams to review content that is reported to violate these standards
- review the majority of the content flagged within 24 hours and remove or disable access to hate speech content, if necessary
- provide regular training to staff
- establish national contact points to communicate with competent national authorities.

The Hate Speech Code is aimed at guiding signatory activities as well as sharing best practices with other internet companies, platforms and social media operators. It is also aimed at engaging in partnerships and training activities with civil society to enlarge the network of trusted reporters, as well as work on promoting independent counter-narratives and educational programs. Signatories also agree to further work on improving feedback to users, education and awareness raising, and being more transparent towards society in general.

One unique aspect of the Hate Speech Code is its approach to compliance and enforcement. Its implementation is evaluated through a regular monitoring

exercise in collaboration with a network of organisations located in the different European Union (EU) countries. Using a commonly agreed methodology, these organisations test how signatories are implementing the commitments in the Hate Speech Code. This includes regularly sending signatories requests to remove content from their online platforms, recording how long it takes them to assess the request, how they respond to the request, and the feedback they receive from the companies.

The EU has suggested that the Hate Speech Code has proven to be an effective policy tool to achieve fast progress by businesses facing a major societal challenge. A 2019 evaluation showed that the Hate Speech Code delivers results: signatories are assessing 89% of flagged content within 24hours and 72% of the content deemed illegal hate speech is removed.

United Kingdom Online Safety Bill (2021)

The United Kingdom (UK) Government's Online Safety Bill states the UK online safety regulator — Ofcom – will set out how companies can fulfil a duty of care to users in codes of practice. These codes will outline the systems, processes and governance that companies need to adopt to fulfil their duty of care.

The UK Government has stated that Ofcom will decide which codes of practice to produce. While there will not be individual codes of practice for each specific harm, there will be specific and individual codes of practice for tackling terrorist use of the internet, and on child sexual exploitation and abuse.

Recognising the need to have specific codes that address this harmful content, the UK Government has released two interim codes. The UK's interim codes take a principles-based approach to encourage industry to adopt best practice in responding to terrorist and violent extremist content (TVEC) and CSEM.

The Interim Code of Practice on Terrorist Content and Activity Online (Interim TVEC Code) provides detailed guidance for companies to help them understand how to mitigate the range of risks arising from online terrorist content and activity in order to protect their users and the public from harm.

Similar to the Hate Speech Code, the Interim TVEC Code encourages industry to focus on the key areas of reporting pathways, timely redress, respect for legal frameworks and industry collaboration. Unlike the Hate Speech Code, the

Interim TVEC Code encourages proactive identification and prevention of content through de-listing/hiding search results.

Of note, the interim codes are designed to align with one another as much as possible to assist companies with understanding and implementing both codes. However, there are some differences as they respond to two different threats.

The Interim code of practice on online child sexual exploitation and abuse (Interim CSEA Code) provides detailed guidance for companies to help them understand and respond to the breadth of CSEA threats, recognising that this threat and the response that it requires will vary depending on the type and nature of the service offered.

Like the Hate Speech and Interim TVEC Codes, there is a focus on the key areas of reporting pathways, timely redress, respect for legal frameworks and industry collaboration. Like the Interim TVEC Code, the Interim CSEM Code encourages proactive identification and prevention of content through de-listing/hiding search results. In addition, there is a strong focus on adopting enhanced safety features and investments in new tools, as well as considering the global threat of CSEM in the design of services.

The interim codes set a high bar for industry, while also recognising that small and medium size enterprises are likely to have less capacity and resources to safeguard their services and therefore may not be able to take the same measures as large companies. Nevertheless, both interim codes recognise that bad actors exploit services of all sizes and varying functions.

Ireland Online Safety and Media Regulation Bill

The Online Safety Media and Regulation (OSMR) Bill, currently before the Irish Parliament, will establish a new Online Safety Commission, under a broader Media Commission, to deal with harmful online content. The OSMR Bill allows the Irish Online Safety Commissioner (Irish Commissioner) to make binding online safety codes that govern standards and practices of online services on a range of topics.

When considering designating services, the Irish Commissioner will take into account the nature and scale of online services, the legal limits of liability and issues of fundamental rights, among other things. The Irish Commissioner can

categorise online services it designates as it sees fit and can designate whole categories of online services.

Given the large range of different kinds of services that the Irish Commissioner may be regulating, the Irish Commissioner will not apply all online safety codes or all aspects of every code to every online service it regulates. Instead, the Commissioner will decide which codes apply to which services it regulates.

In Ireland, it has been identified that the Online Safety Commission will sit under a broader Media Commission, which will be stood up following the dissolution of the Broadcasting Authority of Ireland (BAI). Nevertheless, the BAI's approach to traditional broadcasting codes serves as a useful reference point.

The BAI takes a risk-based approach to enforcing its codes, as outlined in its Compliance and Enforcement Policy.

In this regard, the BAI considers its compliance workplan in the context of:

- the nature, extent and impact of potential non-compliance or breaches
- stakeholder risk including viewer / listener protection and, in particular, children
- reputational risk
- operational (resources) risk
- strategic risk
- financial risk (in particular for investigations and sanctions).

A proportionate and balanced risk-based approach is used to facilitate the effective operation of the BAI. This approach involves the identification and evaluation of risks in accordance with their probability of occurrence (high, moderate and low) and their possible impacts (high, moderate and low) on the achievement of compliance-related and organisational objectives.

German Youth Protection Act – Safety by Design Standard

Section 24a of the German Youth Protection Act has introduced a new Safety by Design Standard for services providers, including to implement pre-emptive protection measures to protect children from content that is harmful to them, or that impairs their development.

Similar to the BAI, the Youth Protection Act takes a risk-based approach, with the higher the risks from a Youth Protection Act perspective, the more robust the pre-emptive measures have to be. For example, the more interactional and

communication risks that are included in a service, the higher the overall risk of such service from a Youth Protection Act perspective: for example, a mobile game that is primarily played by children and that includes in-game chats, in-game purchases, loot boxes and other social network-sharing features will likely have to implement more pre-emptive measures than a mobile game that only has in-game purchases.

Obligations for service providers are similar to other codes in terms of requiring a reporting feature and notice and takedown procedure for harmful content, as well as a requirement to comply with relevant laws. An additional requirement is to provide technical means for age-verification, parental controls and default features that restrict use for minors in consideration of their age. There is also a requirement to include easy-to-find information on third-party counselling and support services.

Another unique feature of the Safety by Design Standard is its approach to compliance and enforcement. The relevant industry body is required to determine whether the standard applies to a particular service. The Safety by Design standard is binding for those services.