# Impact Analysis

**Online Safety (Relevant Electronic Services – Class 1A and 1B Material) Industry Standard 2024**

**Online Safety (Designated Internet Services– Class 1A and 1B Material) Industry Standard 2024**

**June 2024**

## Disclaimer

The material contained in the publication is made available on the understanding that the Commonwealth is not providing professional advice, and that users exercise their own skill and care with respect to its use and seek independent advice if necessary.

The Commonwealth makes no representations or warranties as to the contents or accuracy of the information contained in this publication. To the extent permitted by law, the Commonwealth disclaims liability to any person or organisation in respect of anything done, or omitted to be done, in reliance upon information contained in this publication.

## Creative Commons licence



## Use of the Coat of Arms

The Department of the Prime Minister and Cabinet sets the terms under which the Coat of Arms is used. Please refer to the Commonwealth Coat of Arms – Information and Guidelines publication available at http://www.pmc.gov.au.

## Contact us

Please email us at enquiries@esafety.gov.au with requests and enquiries about permissions.

# Content Warning

## This report contains material that can be confronting and disturbing.

Sometimes words can cause sadness or distress, or trigger traumatic memories for people, particularly survivors of past abuse, violence, or childhood trauma.

For some people, these responses can be overwhelming.

If you need to talk to someone, support is available through redress support services. The following services are available 24 hours a day:

- beyondblue: 1300 224 636
- 1800RESPECT: 1800 737 732
- Lifeline: 13 11 14
- Suicide Call Back Service: 1300 659 467

# About eSafety

The eSafety Commissioner (eSafety) is Australia's independent regulator and educator for online safety. eSafety promotes online safety for all Australians, leads online safety efforts across Australian Government departments and agencies, and works with online safety stakeholders around the world to extend our impact across borders. Established in 2015, our mandate is to make sure Australians have safer and more positive experiences online.

# Acknowledgement

eSafety acknowledges all First Nations people for their continuing care of everything Country encompasses — land, waters, and community. We pay our respects to First Nations people, and to Elders past, present, and future.

# Contents

# Executive Summary

Over recent decades, online services and digital technologies have provided vast benefits to both businesses and users. But the same services are also weaponised to cause online harm.

As internet usage has expanded, Australians are increasingly sharing, storing, accessing, or being exposed to harmful online content such as child sexual abuse material, footage of terrorist acts and extreme violence.

Harmful online content can be seriously damaging, especially for those most at-risk, such as children and young people. The social, emotional, psychological, and physical impact resulting from the production, distribution and consumption of harmful content is felt both immediately and over time. More specifically, the promotion of terrorist acts online can lead to further radicalisation, online incitement of violence can spill over to real-world harm, and the hosting, sharing and proliferation of child sexual abuse material further re-victimises and re-traumatises victim-survivors.

The design, implementation and moderation of online services such as social media services, messaging services and other websites and apps provides a critical role in reducing the risks of these harms.

In response to the growing risks to the Australian community, the Online Safety Act 2021 (the Act) came into effect on 23 January 2022. The objectives in section 3 of the Act are to improve and promote online safety for Australians. The Act built on the pre-existing legislative framework and enhances protections for Australians from online harms, improves industry accountability for the safety of users and enables the eSafety Commissioner to operate as an effective regulator.

The Act and its subordinate instruments apply to online providers operating both within and outside of Australia, where the service they provide can be accessed by persons from Australia (referred to as Australian end-users in this document).

The Act provides for the establishment of new mandatory industry codes and standards for eight sections of the online industry to regulate the most harmful types of online material– class 1 and class 2 material which is material that contains illegal and/or restricted content. This ranges from the most seriously harmful material (such as images and videos showing the sexual abuse of children or acts of terrorism), through to content which should not be accessed by

children (such as simulated sexual activity, detailed nudity, or high impact violence)[1].

In March 2023, eight codes developed by industry associations addressing a subset of class 1 material (class 1A and class 1B material) were submitted to the eSafety Commissioner, following 18 months of development by industry and close discussion with staff of the eSafety Commissioner.

The eSafety Commissioner declined to register two of the eight codes submitted by industry associations:

- the draft *Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Code* which would have covered services that allow end-users to communicate with each other through email, instant messaging, SMS/MMS, chat services or within online games; and
- the draft *Online Safety (Designated Internet Services – Class 1A and Class 1B Material) Code* which would have covered services that allow end-users to access material on the internet such as websites and other online services, but which do not fall within the other categories identified in the Act.

These draft codes were not registered by the eSafety Commissioner because they did not contain appropriate community safeguards, a statutory requirement for registration under sub-section 140(1)(d)(i) of the Act.

In the absence of applicable legal requirements, there is a significant risk of harm to Australians due to the rapid proliferation of high-risk, harmful material on RES and DIS services.

This document outlines the case for, and the estimated impact, of the introduction by the eSafety Commissioner of two new statutory instruments which are referred to throughout this analysis as the standards:

- the *Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024*; and

  the Online Safety (*Designated Internet Services - Class 1A and Class 1B Material) Industry Standard 2024*

[1] Class 1 material is defined in section 106 of the Online Safety Act 2021 (Cth). Class 2 material is defined in section 107 of the Online Safety Act 2021 (Cth).

# Introduction to the Impact Analysis

In accordance with Australian Government policy, any proposals with an expectation of compliance that would result in a more than minor change in behaviour or impact for people, businesses or community organisations are required to complete an Impact Analysis in accordance with the Australian Government Policy Impact Analysis Framework[2].

The Impact Analysis Framework ensures the costs and benefits of new policies are understood from all angles, that decisions are based on evidence and that they best support a stronger economy and guarantee the essentials Australians rely on to prosper. In accordance with the Government Framework, this paper addresses the following seven specified questions as follows:

| Impact Analysis Framework question | Related chapter |
|---|---|
| 1. What is the policy problem you are trying to solve and what data is available? | Chapter 1 |
| 2. What are the objectives, why is the Government intervention needed to achieve them, and how will success be measured? | Chapter 2 |
| 3. What policy options are you considering? | Chapter 3 |
| 4. What is the likely net benefit of each option? | Chapter 4 |
| 5. Who did you consult and how did you incorporate their feedback? | Chapter 5 |
| 6. What is the best option from those you have considered and how will it be implemented? | Chapter 6 |
| 7. How will you evaluate your chosen option against the success metrics? | Chapter 7 |

This document:

- outlines the case for the introduction of Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024; and the Online Safety (Designated Internet Services - Class 1A and Class 1B Material) Industry Standard 2024 (the standards);
- assesses the impacts of the standards; and
- assesses alternative options and their estimated impact, for comparative purposes.

[2] Guidance on Impact Analysis | The Office of Impact Analysis pmc.gov.au) https://oia.pmc.gov.au/resources/guidance-impact-analysis/australian-government-guide-policy-impact-analysis

The standards will operate alongside the six registered industry codes to ensure a set of mandatory, outcomes-based, and technologically neutral requirements for providers of relevant electronic services (RES) and designated electronic services (DIS). Combined with the registered codes, the standards will ensure that each section of the online industry has appropriate community safeguards in place to protect end-users in Australia from harms associated with class 1A and class 1B material.

To date, there has been an inconsistent approach to dealing with material containing illegal and restricted content by RES and DIS providers. The Commissioner's assessment of the draft codes developed by industry for these sections, as well as further information obtained through the exercise of the Commissioner's statutory powers and other research, indicates some companies are not enacting basic safety measures to address the risks to users in Australia from this online material containing this kind of content.

The standards will enhance the protections for Australians from harms caused by class 1A and class 1B material on RES and DIS, enabling the eSafety Commissioner to operate as an effective regulator across these industry sections.

In line with section 13A and section 14 of the Act, a RES or DIS must be accessible to, or delivered to, one or more end-users in Australia to be covered by the Act. The definitions in these sections of the Act are:

| | |
|---|---|
| **RES** | An online service which enables end-users to communicate with one another by email, instant messaging, short message services (SMS), multimedia message services (MMS) or chat services, as well as services that enable end-users to play online games with each other, and online dating services. |
| **DIS** | An online service which allows end-users to access material using an internet carriage service, or which delivers material to persons who have equipment appropriate for receiving that material, where the delivery is by means of an internet carriage service but excludes social media, RES, and other identified services. This category includes many apps and websites, as well as online storage services which are used by end-users to upload, store, and manage their files including photos and other media. |

This document examines three regulatory options and assesses their potential impact for the point of comparison.

**Option 1 (maintain the status-quo)** would see no enforceable regulatory requirements on RES and DIS sections of the online industry to have systems and processes in place to deal with class 1A and class 1B material[3]. RES and DIS providers would not be subject to any legal regulatory requirements to proactively address the serious harms caused by class 1A and class 1B material. Users would continue to be reliant on voluntary steps made by RES and DIS, which have been insufficient to address this illegal and harmful content to date and eSafety's powers to direct the take down of individual pieces of content.

**Option 2 (industry co-regulation)** would require the development of RES and DIS industry codes for class 1A and 1B online material which are able to be registered by the eSafety Commissioner. While draft industry codes have been developed, the Commissioner declined to register these in May 2023 as they did not provide adequate community safeguards. While implementation of the draft codes would provide some additional protection to Australian end-users from class 1A and class 1B material on RES and DIS services, these codes were rejected because they did not provide appropriate community safeguards. It is not expected that further consultation with industry would result in appropriate RES and DIS codes and it would only further delay protecting Australian's online.

**Option 3 (direct regulation)** is that the eSafety Commissioner register the standards, putting in place proactive obligations on these services. Registration of the standards ensures that appropriate community safeguards to protect Australian end-users against class 1A and class 1B material are in place across these industry sections, consistent with the objectives of the Act in section 3 and the eSafety Commissioner's statutory functions in section 27 of the Act.

[3] RES and DIS will continue to be required to comply with notices issued by eSafety under the Online Safety Act 2021 (the Act) to remove content (after it has been surfaced) or to provide information to eSafety requested pursuant to a statutory notice, in connection with the Online Safety (Basic Online Safety Expectations) Determination 2022 (the BOSE).

Consultation on draft standards provided the eSafety Commissioner with feedback from a wide range of stakeholders include service providers, industry associations and civil society. The draft standards were amended and finalised on consideration of the feedback received during consultation. Changes were made in several key areas including the test to determine which code or standard is applicable to a certain service provider, an additional exception to address security vulnerability, clarifying the detection and removal of pro-terror materials obligation and the generative AI categories under the DIS standard.

Online harms have a profound impact on Australians. A recent report by the Australian Institute of Criminology on the online sexual exploitation of children highlights that '[child sexual abuse material] is a significant societal problem that causes and perpetuates long-lasting harm to victims, who are both directly sexually abused and repeatedly re-victimised through the ongoing distribution and accessing of [child sexual abuse material} long after the abuse occurs' (Australian Institute of Criminology, 2022). Victim-survivors and their families are also retraumatised by inadequate responses from technology companies in requests to remove material depicting the sexual abuse or exploitation of their child. Exposure to pro-terror and extreme violence material also has the potential to both cause individuals harm, as well as potentially impacting all Australians through the radicalisation of at-risk individuals leading to an increase in real-world violence.

It is critical that RES and DIS - which include high risk services for accessing, sharing, and storing class 1A and class 1B material such as some pornography websites, chat, messaging services and photo storage services - have robust and enforceable community safeguards proportionate to the risk of harm from class 1A and class 1B material on these services. For this reason, option 3 is the preferred option as it most effectively promotes and improves online safety for all Australians.

# 1. What is the policy problem you are trying to solve and what data is available?

**This chapter outlines the policy problem, which is to protect Australians from the harms caused by the production, transmission, and consumption of class 1A and class 1B material on RES and DIS. Evidence on the harm caused by this material, and its prevalence on RES and DIS, is examined along with the role of RES and DIS in the creation, distribution and storage of such material.**

## 1.1. Seriously harmful material is shared, stored, accessed, and generated on RES and DIS

Access to the internet and technological developments continue to provide new opportunities for Australians to engage and connect with each other, and to access and share material online.

In Australia, 97% of households with children aged under 15 years had access to the internet, as at 2016-17 (Autralian Bureau of Statistics, 2018). Consumer technologies that allow access to the Internet have become ubiquitous within Australian households. According to eSafety research (2018), 91% of parents with pre-schoolers report that their children connect to the internet through a smartphone (eSafety Commissioner, 2018) and 81% of parents with pre-schoolers in Australia (children aged 2-5) say their children use the internet (eSafety Commissioner, 2021). A University of NSW (2021) study found that according to parents and grandparents of children aged 5-17 who were surveyed 'more than 4 in 5 children own at least one screen-based device... and children own, on average, three digital devices at home' (Graham & Sahlberg, 2021).

More time spent on screen-based activities and internet-connected devices has increased the likelihood of exposure to online harm for all Australians, but particularly children. The Australian Centre to Counter Child Exploitation (ACCCE) has stated that the increase in young people accessing the internet has seen a corresponding upward trend in cases of online child sexual exploitation (Australian Centre to Counter Child Exploitation, 2021).

The democratisation of powerful technologies at relatively low cost, and without embedded safeguards, has made it much easier for to create, distribute, and consume online material containing illegal and restricted content. The worst of this material is categorised as class 1A material including child sexual exploitation material, pro-terror material and extreme crime and violence material. Harmful behaviours that lead to the creation of this kind of material include online grooming of children to sexually abuse them, or to expose them to extremist content and radicalise them.

The advancement of generative AI also further facilitates harms, as recent studies by the Stanford Internet Observatory and Thorn (Thiel, Stroebel, & Portnoff, 2023) highlight the rapidly advancing threat of the production of highly realistic child sexual abuse material using generative AI models.

Given what we know about the scale of this problem, it is critical to ensure safeguards are introduced to reduce the risk of harm arising from class 1A and class 1B material on RES and DIS including child sexual abuse material, pro-terror material and extreme crime and violence material.

## 1.2. The limitations of our current regulatory framework

The *Online Safety Bill 2021* (Cth) was passed by the Australian Parliament on 23 July 2021, with the Act coming into effect on 23 January 2022. Part 9, Division 7 of the Act provides for the establishment of new mandatory industry codes and standards for eight sections[4] of the online industry.

The Act provides for industry bodies to develop codes and for eSafety to register the codes if they meet the statutory requirements. The codes become enforceable when registered by the eSafety Commissioner. If a draft code does not meet the statutory requirements, the eSafety Commissioner is able to determine an industry standard for that section of the online industry.

On 31 May, the eSafety Commissioner determined that the draft RES and DIS codes submitted by industry associations did not provide appropriate community safeguards. Without any further regulation RES and DIS providers would have a

[4] Six industry codes came into effect in December 2023 for social media services, internet carriage services (also known as internet service providers), equipment providers, app distribution services and hosting services. The industry code for internet search engines will come into effect 12 March 2024.

significantly lower level of regulation compared to those industry sections subject to a code.

## 1.3. What is class 1A and class 1B material?

Class 1 and class 2 material are defined under the Act by reference to Australia's National Classification Scheme[5].

- **Class 1 material** (defined in section 106 of the Act) – is material that is or would likely be refused classification under the National Classification Scheme.
- **Class 2 material** (defined in section 107 of the Act) is material that is, or would likely be, classified as either X18+ or R18+ under the National Classification Scheme (because it is considered inappropriate for public access and/or for children and young people under 18 years old).

To facilitate implementation of the Act, eSafety developed subcategories of class 1 material and asked industry to take a two-phased approach to developing industry codes. The purpose of this was to prioritise the implementation of measures to prevent and reduce the most harmful online material. Industry and stakeholders supported this approach.

Phase 1 implementation deals with the most harmful material, that is material which is described as **class 1A or class 1B material**.

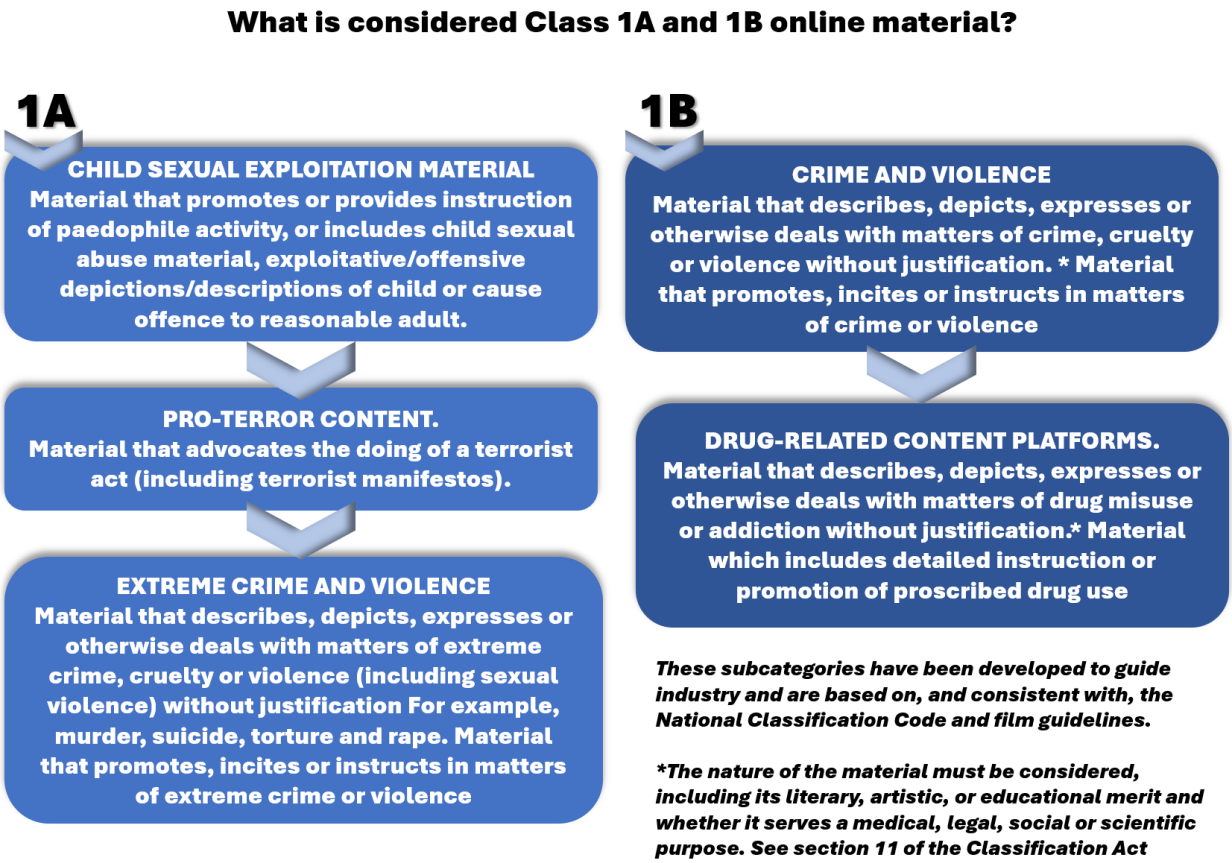
These are sub-categories of class 1 material which were developed by eSafety in recognition that they constitute the most harmful material and should be dealt with as a priority.

Class 1A and Class 1B material is summarised in Figure 1 overleaf.

Subsequent industry codes (or if required, industry standards) will be developed to address class 2 (restricted R18+/X18+) material, such as online pornography and other high impact material as well as material identified by eSafety as class 1C material (certain fetish pornography falling within the definition of class 1 material)[6].

[5] A cooperative arrangement between the Australian Government and state and territory governments for the classification of films, computer games and certain publications. For further information visit the Australian Classification website at www.classification.gov.au.
[6] See page 23 of the eSafety Position Paper

Figure 1




## 1.4. What harms does this problem cause?

Harms attributable to class 1A and 1B material available online can be grouped as follows:

- Harms arising from production of the material– for example, where a perpetrator grooms, coerces, or forces a child into the production of content, or where coerced sexual activity or abuse of a child is recorded or, in the case of pro-terror content, when a perpetrator carries out terrorist activity which is recorded to distribute for propaganda purposes.
- Harms arising from distribution of the material– for example, where abusive material is posted, reshared or live-streamed online, which can compound the trauma experienced by survivors tortured, sexually abused, and harmed in the production of content. Victim-survivors of terrorist activity and their families are similarly harmed when footage of an attack is distributed and remains available online.

- Harms arising from the consumption of the material– for example, where a person's behaviour, emotions, mental health, attitudes, or perceptions are negatively impacted because of access or exposure to harmful content. For example:
    - a recent survey of viewers of child sexual abuse material on the dark web found that one third of respondents attempted to directly contact a child following viewing child sexual abuse material online (Insoll, Ovaska, & Nurmi, 2022);
    - individuals exposed to, imitating and internalising extremist beliefs and attitudes via the internet can be understood as undergoing online radicalisation. Such radicalised individuals are seen as at an increased risk of committing offences, such as violent acts of terrorism (Binder & Kenyon, 2022).

The production, distribution, and storage of class 1A and class 1B material and the consequent consumption of it causes serious and long-term physical, psychological, and financial damage to victim-survivors, to their families and communities, and to the Australian economy.

### 1.4.1. Harms from online child sexual abuse material

Many children who are the subject of online child sexual abuse may suffer ongoing harms from the sexual abuse or exploitation itself, and from the repeated sharing and viewing of the abuse material (Gewirtz-Meydan, Walsh, Wolak, & Finkelhor, 2018) (Joleby, Lunde, Landstrom, & Jonsso, 2020).

As highlighted by the International Justice Mission in its testimony before a 2023 US Congressional hearing on child exploitation:

> 'behind every livestream is a real child, suffering serious emotional and physical trauma… there is no end to their continued exploitation and the invasion of their privacy, as offenders share and trade images and videos of child abuse in encrypted messaging apps and online' (International Justice Mission, 2023).

Ruby* who was sexually abused in livestreams as a 16-year-old, recalls how the abuse impacted her life:

'While doing every disgusting show [in front of the computer camera with the customer], I lost every bit of my self-esteem to the point where I felt disgusted with myself as well. It's like being trapped in a dark room without any rays of light at all. There's no point in living at all' **(WeProtect Global Alliance, 2023)**.

Online harms have a profound impact on victim-survivors. A recent report by the Australian Institute of Criminology (AIC) on the online sexual exploitation of children highlights that child sexual abuse material 'is a significant societal problem that causes and perpetuates long-lasting harm to victims, who are both directly sexually abused and repeatedly re-victimised through the ongoing distribution and accessing of [child sexual abuse material] long after the abuse occurs' (Gewirtz-Meydan, Walsh, Wolak, & Finkelhor, 2018).

This was also highlighted in research examining the impacts of online child sexual abuse and exploitation, where it was found that victim-survivors reported experiencing psychological trauma, anxiety, depression and self-harming or suicidal behaviour because of the abuse. They also reported self-blame, trust issues, impaired relationships, and difficulties at school. The impacts were felt by victim-survivors into adulthood, affecting family and intimate relationships (WeProtect Global Alliance, 2023) (Australian Institute of Criminology, 2022).

The development of generative AI has created new harms and risks. Images of real children can be manipulated to create sexualised depictions of them. Even where generated material is not based on actual children, it causes harm to victim-survivors of child sexual abuse. A recent report by Stanford Internet Observatory and Thorn (Thiel, Stroebel, & Portnoff, 2023) highlighted that child sexual abuse material is being generated that is almost indistinguishable from actual images. This presents several challenges as 'in a scenario where highly realistic computer-generated CSAM (CG-CSAM) becomes highly prevalent online, the ability for NGOs and law enforcement to investigate and prosecute CSAM cases may be severely hindered'.

**Responses from parents of online child sexual abuse victims highlighting the broader impacts:**

'I already was super protective: I home-schooled, limited online time, used family search safety utilities, DNS blockers, and buddy system with my kids. I feel like a failure, and I regret getting married and having a family' **(Canadian Centre for Child Protection Inc, 2023)**.

'We worry it will be shared, used for people to extort more images, used for bullying, accessible if she applies for a job, education or gets into a relationship, even a healthy relationship with someone. We worry it will be shown to other kids to make it seem normal and further child sexual abuse' **(Canadian Centre for Child Protection Inc, 2023)**.

Victim-survivors and their families are also retraumatised by inadequate responses from technology companies in requests to remove material depicting the sexual abuse or exploitation of their child.

'When parents asked technology companies to take down the abuse imagery or other harmful content, companies rarely complied. Some simply refused, while others said they would only remove the material if parents provided them with information about the child in the abuse imagery... technology companies often have few to no barriers for the uploading of [child sexual abuse material], while putting up many barriers throughout the system for parents seeking to have the imagery removed' **(Canadian Centre for Child Protection Inc, 2023).**

The following quotes from survivors of child sexual exploitation underscore the deep and prolonged harm of child sexual abuse material.

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

**(Canadian Centre for Child Protection Inc, 2017)**

### 1.4.2. Harms from online pro-terror and extreme violence material

Exposure to pro-terror and extreme violence material has the potential to cause individuals harm as well as potentially impacting all Australians through the radicalisation of at-risk individuals leading to an increase in real-world violence. Young people are particularly vulnerable to harms from pro-terror and extremist material (Commission for Countering Extremism, 2020).

A year before he attacked two mosques in Christchurch, New Zealand, the individual responsible for the attack posted publicly online about his plans (Ko tō tātou kāinga tēnei. Report: Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019. December 2020). In their investigation of right-wing websites and whether they were an important factor in the individual's radicalisation, researchers found that he had been posting anonymously on the online message board 4chan up to four years prior to the attacks about his desire to attack persons of colour in significant locations including places of worship and concluded that the 4chan community was crucial in the individual's radicalisation. His final post on the imageboard 8chan, but also intended for 4chan, being 'It's been a long ride [...] you are all top blokes and the best bunch of cobbers a man could ask for' (Wilson, C., et al The Conversation. 21 February 2024).

The Commission for Countering Extremism highlights the six main harms resulting from the consumption, production, and distribution of this material as (Commission for Countering Extremism, 2020):

- social division and intolerance
- crime, violence, and harassment
- mental health and wellbeing
- censorship and restriction of freedom
- delegitimising authority/undermining democracy
- economic harms

Perpetrators use extreme violence material on RES and DIS to amplify and promote their terrorist agendas and violent crimes. The Australian Security Intelligence Organisation (ASIO) has highlighted that the internet plays an important role in the radicalisation, recruitment, indoctrination and training of future violent extremists and terrorists. Radicalisation of individuals can occur

both face-to-face and through a virtual environment online where an individual may become part of an online community of people who share their hateful views and ideologies (Australian Security Intelligence Organisation, 2019) (Australian Government Attorney-General's Department, 2015).

## 1.5. The scale of class 1A and class 1B material available online

The digital environment has become an enabler for the production, distribution, and storage of material containing illegal and restricted online content, including child sexual abuse material and pro-terror material. The volume of this harmful content online is significant and continues to increase in scale, severity, and complexity.

This is evidenced by increased reports across a range of reporting schemes under the Act. During 2022-23, through the Act's Online Content Scheme eSafety received 11,636 complaints concerning 33,129 Uniform Resource Locators (otherwise referred to as URLs[7]), with 87% related to child sexual abuse, child abuse or paedophile activity. This is a 110% increase from 2021-22 (Australian Communications and Media Authority and eSafety Commissioner, 2023). The Australian Centre to Counter Child Exploitation also received 33,114 reports of online child sexual exploitation in 2021, almost double the number received in 2018 (which was 17,400) (Australian Centre to Counter Child Exploitation, 2021).

Internationally, the Internet Watch Foundation (2022) identified a 64% increase in URLs containing or advertising child sexual abuse material in 2021. The US based National Centre for Missing and Exploited Children (NCMEC 2024)[8] received 36.2 million reports of child sexual exploitation and abuse in 2023, including 54.8 million images and 49.5 million videos from tech companies. Although these are the total reports from all online services, they reflect a significant number of notifications from RES and DIS. For example, WhatsApp, which is a RES, made

[7] A URL is the address of a given unique resource on the Web. In theory, each valid URL points to a unique resource. Such resources can be an HTML page, a CSS document, an image etc.
[8] NCMEC is the US's national clearinghouse for reporting CSAM materials online in the US and operates a CyberTipline which provides an online mechanism for members of the public and electronic service providers to report incidents of suspected child sexual exploitation. NCMEC then makes these reports available to law enforcement agencies around the globe.

around 1.4 million reports of child sexual abuse material to NCMEC in 2023. WhatsApp had also previously reported that it acts against hundreds of thousands of accounts each month for suspected sharing of child exploitation imagery (WhatsApp, n.d.). NCMEC reports also demonstrate a significant increase in financial sextortion schemes targeting teens and children with a 7200% increase from 2021 to 2020 (WeProtect Global Alliance, 2023). NetClean's Covid-19 Impact Report 2020 also identified a sharp increase in 'online enticement' (i.e., grooming, or sexual extortion), with cases doubling over a 12-month period between 2019-2020 (NetClean, 2021).

While these reporting rates demonstrate significant increases in the production and distribution of child sexual abuse and exploitation material, the true scale of this abuse online is likely much greater than what is being captured, as most incidents are not reported (WeProtect Global Alliance, 2023).

These statistics represent only the tip of the iceberg.

Terrorist and extremist groups have also exploited pandemic conditions to radicalise, incite and amplify hate and grow support for violent activities (Commission for Countering Extremism, 2020). According to the Institute of Strategic Dialogue, the pandemic 'created a febrile environment for radicalisation, by ensuring that millions of people have spent more time online [and] in an environment of heightened anxiety the situation [was] an easy one for extremists to capitalise on' (Hart, Davey, Maharasingam-Shah, & Gallagher, 2021).

While a significant proportion of this extremist and pro-terror material is increasingly being circulated and distributed through social media platforms, 'de-platforming' of groups and individuals has also pushed them towards the use of RES such as private messaging platforms (Commission for Countering Extremism, 2020). The distribution of terrorist and violent extremist material online has been demonstrated to be a crucial component of terrorist and extremist groups' radicalisation operations (Llanos, 2022). In the United Kingdom, research has found that all terrorist attacks carried out since 2017 have had an online element (Llanos, 2022).

There is extensive evidence that the generation, distribution, and consumption of class 1A via RES and DIS services is systemic and increasing (OECD, 2023).

Australian Institute of Criminology research found that:

'The platforms with the highest user bases are actively detecting and removing [child sexual abuse material]. However, some are less transparent than others about the methods they use to prevent, detect and remove [child sexual abuse material], omitting key information that is crucial for future best practice in reducing [child sexual abuse material] offending. Further, the adoption of end-to-end encryption by platforms that detect and remove large amounts of [child sexual abuse material] from their platforms will likely provide a haven for [child sexual abuse material] offenders.'

**(Teunissen & Napier, Child sexual abuse material and end-to-end encryption on social media platforms: An overview, 2022)**

Live streaming services are widely known to be vehicles for the online exploitation of children. The livestream may be recorded, and stored on cloud services from where it can be disseminated via websites or messaging/chat services. These services are used to enable the live streaming of sexual abuse of children either lured or forced into sexual acts which are recorded, and the abuse then broadcast to other offenders (WeProtect Global Alliance, 2023). The live streaming of child abuse occurs disproportionately from low-income countries such as South-East Asia into high-income countries (Napier, Teunissen, & Boxall, How do child sexual abuse live streaming offenders access victims?).

The AFP (2021) has found that Australian children are being targeted online and coerced into performing livestreamed sexual acts. Perpetrators record the videos, share them online, and/or extort victims into producing even more graphic content. The AFP considers this practice, known as 'capping' (short for capturing), to be one of the fastest growing trends in online child sexual abuse. Law enforcement agencies internationally are also reporting that offenders are recording livestreams to obtain content with which to 'sextort' their victims into further acts (Napier & Teunissen, 2023). Accessing child sexual abuse material via live streaming services continues to increase, with reports suggesting the global demand is high (Australian Institute of Criminology, 2022).

In AFP Operation Molto (Australian Federal Police, 2022) and more recently NSW Police Force Strike Force Packer (NSW Police Force, 2024) organised offender groups have been found producing online child sexual abuse material in Australia for global distribution. Operation Molto resulted in the charging of more than 100 Australians with over 1,000 child abuse-related offences. Coordinated by the AFP-led Australian Centre to Counter Child Exploitation (ACCCE) and working together

with police from every state and territory in Australia, police executed 158 search warrants in Australia, charging 121 men with 1,248 offences and removing 51 Australian children from harm. Operation Molto commenced in 2019, when the ACCCE received intelligence from New Zealand's Te Tari Taiwhenua Department of Internal Affairs showing thousands of offenders were using a cloud storage platform to share abhorrent child material abuse online. The multinational law enforcement effort resulted in 153 children being removed from harm, including:

- 51 children in Australia
- 79 children in the United Kingdom
- 12 children in Canada
- 6 children in New Zealand
- 4 children in the United States, and
- 1 child in Europe.

Some of the alleged offenders in Australia were producing their own child abuse material and were found in possession of material that was produced by a man arrested by the AFP in 2015 under Operation Niro, which resulted in the dismantling of an international organised paedophile syndicate. This material was classified as the most abhorrent produced (Australian Federal Police, 2022).

In March 2024, NSW Police Force Sex Crimes Squad detectives announced they had charged a ninth man over his involvement in an international child abuse ring. Strike Force Packer was established in March 2023 by officers attached to the Child Exploitation Internet Unit to investigate an international child abuse ring who were allegedly sharing and viewing child abuse material in online video conferences. The group included national and international members, and NSW Police shared information with Queensland Police, WA Police, Victoria Police, AFP, and the FBI. In NSW alone, 9 men who were identified as taking part in the online group were charged with over 70 offences by NSW Police (NSW Police Force, 2024).

While the exact proportion of live streaming of child sexual abuse that is recorded is unknown, there is a market for this. A recent scoping review conducted by academics on the live streaming of child sexual abuse identified RES such as Skype and Facebook messenger as well-established platforms used to initiate and facilitate live streamed child sexual abuse (Drejer, Riegler, Halvorsen, Baugerud, & Johnson, 2023). The authors of this study recommended that 'policymakers must be made aware of the rising threat livestreaming services present to society and

its children. Policymakers should focus on holding companies accountable for the platforms they provide' (Drejer, Riegler, Halvorsen, Baugerud, & Johnson, 2023).

## 1.6. RES and DIS and the creation, distribution and storage of class 1A and class 1B material

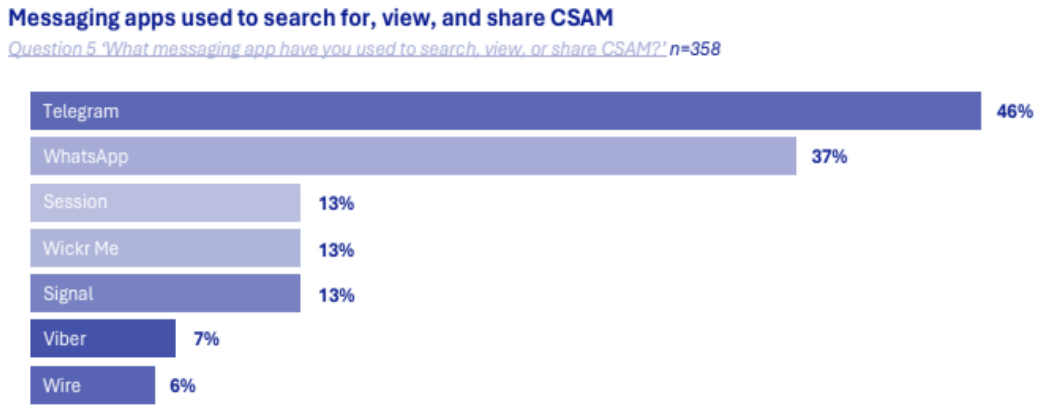
There is significant evidence that RES such as email, instant messaging, videoconferencing, dating and gaming services, and DIS such as online file storage services as well as other websites and apps (including for example, pornographic websites, terror sites or image generators) are being used to create, distribute, access and/or store class 1A and class 1B material.

### 1.6.1. Child sexual abuse material on messaging and gaming services (RES)

Messaging services, including private and end-to-end encrypted messaging services, (which are RES) are used by offenders to network and exchange child sexual abuse material (ECPAT International; INTERPOL; UNICEF, 2024). End-to-end encryption can be an important measure for protecting sensitive information, however it can also create significant risks to the safety and ongoing privacy of children. Messaging services and peer-to-peer networks that use end-to-end encryption create private environments that are preferred by many perpetrators due to the lower risk of detection, which means they can be used as a mechanism to groom children and enable perpetrators to share abuse material and methodologies (WeProtect Global Alliance, 2023).

Recent research by the non-profit organisation Protect Children found that 29% of its survey respondents used a messaging application to search for, view, or share child sexual abuse material, with 37% of respondents stating 'that they established the first contact with a child via a messenger, mostly via end-to-end encryption messengers' (Suojellaan Lapsia, Protect Children ry., 2024).

**Figure 2 Messaging apps used to search for, view and share CSAM** (Suojellaan Lapsia, Protect Children ry., 2024)




## Case study 1

**Reports of an international network of offenders using in-game communications and messaging platforms to access children and extort them to sexually exploit and grievously harm themselves**

**Trigger Warning: This contains content that can be confronting and disturbing.**

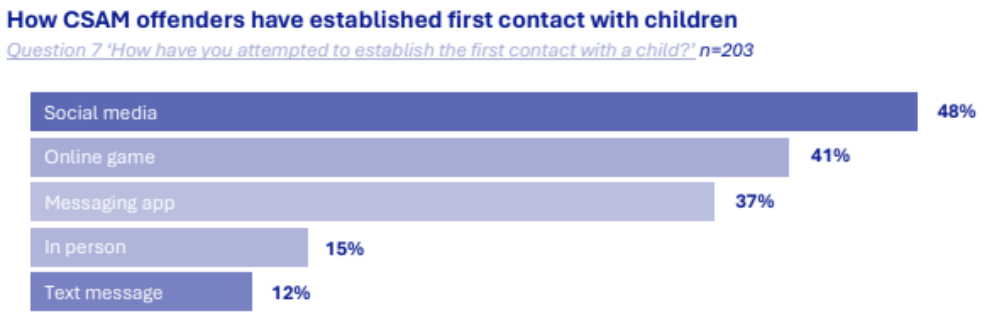
In March 2024, a consortium including Der Spiegel, Recorder, The Washington Post, and WIRED reportedly uncovered an international network of violent predators ('764' extortion network) using the messenger platform Telegram and gaming platforms Minecraft and Roblox to access children in multiple countries and extort them to sexually exploit and grievously harm themselves, including being pushed to suicide.

The network of predators is reported to have coerced children into sexual abuse and self-harm (including carving the abuser's online alias into their skin"). According to the reports, victims carried out violence activities on family members, harming animals, and that in some extreme instances the coercion led to suicide.

**(Winston, 2024)**

Gaming platforms and private messaging services (both types of RES) are used by offenders to initiate contact with children and groom them. The typical modus operandi involves perpetrators targeting children on social media and gaming platforms then moving interactions to a private messaging platform where there is lower risk of detection (WeProtect Global Alliance, 2023).

**Figure 3 - How CSAM offenders have established first contact with children**
(Suojellaan Lapsia, Protect Children ry., 2024)




Online gaming environments have rapidly innovated and expanded and most young people in Australia now regularly play games. In research conducted by eSafety, 89% of the young people surveyed had played online games in the past year, with most (66%) playing for more than 6 hours per week. Four out of 5 (79%) young gamers had played with others online, including 2 in 5 (40%) who had played with people they didn't already know offline and 1 in 4 (26%) who had communicated while gaming with players they didn't already know offline (eSafety Commissioner, 2024).

This increased participation of young people in online gaming has increased the risk of exposure of Australian children to predators who engage in online grooming and other harmful behaviours, such as 'offenders [who] use game-based incentives, like in-game currency, to groom children into sending them child abuse material' (Australian Federal Police, 2023).

By its very nature, online gaming normalises communications with strangers. In-game communications with other players is a core aspect of the activity, leading to children being less suspicious of strangers and less attuned to danger (Suojellaan Lapsia, Protect Children ry., 2024). Recent research by (WeProtect Global Alliance, 2023) has shown that '45 minutes is the average time for a high-risk child grooming situation to develop in social gaming environments, but this can be as quick as 19 seconds."

> In online games, many adults wanted nude photos and tried to pressure me into taking them. Many bets on sexual role-plays, for which I received goods in the games as a reward. - Survivor of childhood sexual violence **(Suojellaan Lapsia, Protect Children ry., 2024)**

### 1.6.2. Child sexual abuse material on file/photo sharing and cloud services (DIS)

End-user managed online file and image storage services are examples of DIS. An end-user is the consumer of a service – in this case a person who uses an online file and photo storage service. File and image storage services are used by perpetrators to store and share child sexual exploitation material. In 2021 the UK-based child safety non-profit organisation Internet Watch Foundation found that image storage websites which allow file or photo sharing were the predominate source of child sexual exploitation images detected (Internet Watch Foundation, 2022).

File and image storage services can provide a hosting environment that makes the distribution of child sexual abuse material reasonably simple. In July 2023, an Australian man was arrested for uploading hundreds of photos and videos of children being abused to a cloud-based storage account. The man was charged under the *Criminal Code Act 1995* (Cth) for using a carriage service to access, transmit and possess or control child abuse material and was imprisoned for almost 3 years (Australian Federal Police, 2023).

### 1.6.3. Pro-terror and extreme crime/violence material on RES and DIS

Pro-terror material includes any material that directly or indirectly counsels, promotes, encourages, instructs, or urges a terrorist act[9]. Extreme crime and violence material includes content that shows, describes, promotes, incites, or instructs people in violent crimes including terrorist acts, kidnapping with violence or threats of violence, murder, attempted murder, rape, torture, and suicide (eSafety Commissioner, 2021).

Where possible, violent extremists will often look to livestream terrorist attacks and recording of the livestream can lead to viral dissemination across the internet. Once pro-terror material circulates online, it is nearly impossible to identify and remove all instances and its continued availability incites further radicalisation and terrorist activity. For example, eSafety continues to receive reports of the

[9] Classification Act 1995 (Cth) s 9A(2).

video footage from the 2019 Christchurch Mosque terrorist attack. eSafety can use its powers under Part 9 of the Act to direct the removal of individual instances of this material, when reported to us.

AI generated pro-terror material also has the potential to contribute to insidious and cumulative harms by influencing public perceptions and values, including towards extremist ideologies (eSafety Commissioner, 2023). A recent report by Tech Against Terrorism (2023) identified terrorist and violent extremist actors engaging with generative AI to augment current practices of creating and disseminating terrorist and violent extremist propaganda.

Gaming platforms (which fall within the RES section) are used by terrorists to radicalise and recruit, and to propagate their ideology (Tech Against Terrorism, 2022). Gaming platform chat functions have also been used to communicate and plan, as well as live-stream, attacks. The AFP has reported that extremists may use popular online chats and other forums such as gaming platforms to recruit Australian children (Schultz, 2023).

So-called ‘gore sites’ – websites which specialise in sharing graphic and disturbing violent material (which fall within the DIS section) – are known vectors for the dissemination of pro-terror and violent extremist material (Hardy & Stewart, 2023). There is evidence ‘gore sites’ serve as digital hubs for the sharing of real-life killings, torture, and other forms of violence, both to a niche audience searching for graphic and disturbing material, and to a secondary audience in the form of violent extremist groups (Hardy & Stewart, 2023).

File storage services (which are a type of DIS) are also used by extremists to store pro-terror and violent extremist material and aggregate information, such as lists of URLs to allow easy access to additional content (Tech Against Terrorism, 2022). The Counter Extremism Project, a not-for-profit international policy organisation formed to combat the growing threat from extremist ideologies, reports that terrorists exploit cloud storage providers to share and stream content, radicalise, and incite violence (Counter Extremism Project, 2018).

Terrorist and violent extremist groups have also been detected operating websites to provide a centralised mechanism to disseminate propaganda, network, recruit and generate funds online (Tech Against Terrorism, 2021).

DIS communications applications that deploy end-to-end encryption on parts of their services, such as Telegram, are reported to be widely used by well-known terrorist and extremist groups to recruit new members and incite violence. Telegram has been described as having a 'flexible interface [which] enables extremists to do everything from self-promotion, brand development and propaganda dissemination' and avoid law enforcement detection (Counter Extremism Project, 2024).

Figure 4 below draws on work by the Global Internet Forum to Counter Terrorism (a non-government organisation established by industry to prevent terrorists and violent extremists from exploiting digital platforms) to show how different types of RES and DIS services are used for the distribution of terrorist and violent extremist material.

**Figure 4: How examples of RES and DIS are used for terror and violent extremist material distribution – taken from the GIFCT Technical Approaches Working Group**

(Tech Against Terrorism, 2021).

## 1.6.4. Class 1B crime and violence material and drug-related material on RES and DIS

Class 1B crime and violence material refers to material that describes, depicts, expresses, or otherwise deals with matters of crime, cruelty or violence without justification, and material that promotes, incites, or instructs in matters of crime or violence.

Class 1B drug-related material refers to any material that describes, depicts, expresses, or otherwise deals with matters of drug misuse or addiction without justification, or which instructs or promotes drug use.

### 1.6.4.1. Crime and violence material on RES and DIS

There is limited research on the volume of production or distribution of class 1B crime and violence material specifically on RES and DIS platforms. Research and studies focus more on the connection of class 1A extreme crime and violence to pro-terror and extremist radicalisation, rather than the class 1B crime and violence material.

While not explicit to RES or DIS platforms, recent eSafety research demonstrates that young people in Australia are exposed to violent and crime-related material online. A 2022 study found that over a third (37%) of the Australian young people aged 14-17 years who were surveyed said they had seen gory or violent images or videos, and one in five (23%) young people aged 14-17 surveyed said they had seen violent sexual images or videos on websites or online discussions (eSafety Commissioner, 2022).

More recent research by eSafety found that 6% of young Australians aged 13-17 years who play games online had seen other players show, share, or talk about things that are illegal in real life, and 3% had seen others sharing violent pictures or videos including of real people being hurt or killed (eSafety Commissioner, 2024).

**Case Study 2**

**Children encountering violent online content**

In March 2024, the United Kingdom online safety regulator, Ofcom, released a report that explored the pathways through which children encounter violent content online. Although conducted in the United Kingdom, the study examined global platforms including services that would be considered RES or DIS in Australia and covered by the standards. Ofcom's key findings included:

- Children described encountering violent content as 'unavoidable.'
- Children had first seen violent content in primary school.
- Children had seen a wide range of violent content, mostly via social media, video sharing services and messaging services. Children also mentioned seeing violent content on online gaming and chat room forums.
- Many children were encountering violent content without seeking it out.
- Professionals and children think platforms have a responsibility to protect children from violent content.

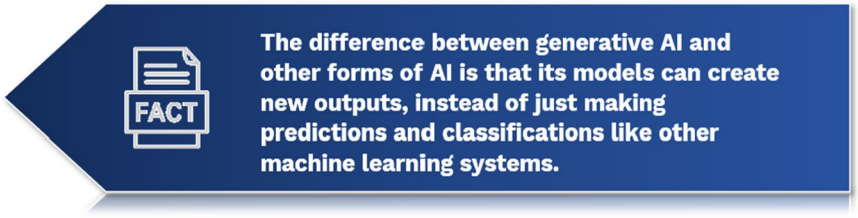
**Question: 'What kinds of violent content do children see online?'**
**Answer: 'Fights, weapons, pain, promotion of gangs (roadmen, clothing, seeing groups)' – Boy, West Yorkshire, 15**

**(Ofcom, 2024)**

#### 1.6.4.2. Drug-related material on RES and DIS

There is limited research and data as to the nature and volume of class 1B drug-related material specifically on RES and DIS platforms. Research and studies into drug-related harms focus primarily on exposure through social media platforms or the internet generally. eSafety research has found that over a third (37%) of young people aged 14-17 surveyed said they had seen websites or online discussions where people talk about or show their experiences of taking drugs (eSafety Commissioner, 2022).

### 1.6.5. AI-generated class 1A and 1B material on DIS




'Generative AI' is a term used to describe the process of using machine learning to create digital content such as new text, images, audio, video, and multimodal simulations of experiences. The difference between generative AI and other forms of AI or machine learning (which has been in use for much longer) is that its models can create new outputs, instead of just making predictions and classifications like other machine learning systems (eSafety Commissioner, 2023). Some examples of user-facing generative AI services include text-based chatbots, or programs designed to simulate conversations with humans. (eSafety Commissioner, 2023). Many online providers offering generative AI services, will be DIS.

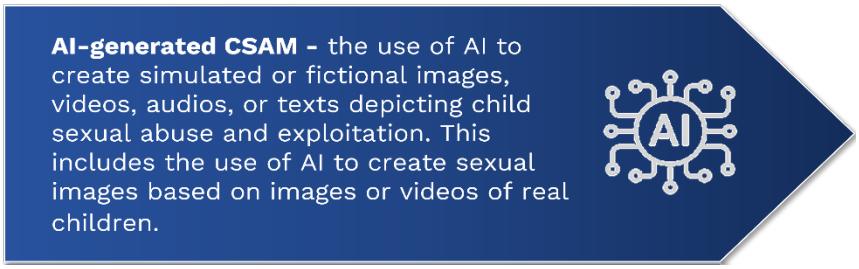
While generative AI can be used as an effective tool to enhance online safety, for example by detecting and moderating harmful online material, it can also be misused to create high-impact child sexual abuse material, and pro-terror and extreme violence material.

Research (Thiel, Stroebel, & Portnoff, 2023) suggests that generative AI can and is being used to create the following types of harmful material, which can be generated via a DIS and stored, distributed, or accessed via a RES or DIS:

- highly realistic synthetic imagery depicting child sexual exploitation and abuse, and pro-terror material (Thiel, Stroebel, & Portnoff, 2023). Perpetrators train generative AI models on existing child sexual abuse material to generate further material of victims; and
- authentic-seeming content for the purpose of bullying, abusing, or manipulating a target – including, but not limited to, grooming children for exploitation (eSafety Commissioner, 2023).

Multi-modal capabilities that analyse social media posts, online interactions, and other data sources can also be weaponised by terrorist groups and violent

extremists to create tailored propaganda, to radicalise and target specific individuals for recruitment, and to incite violence (eSafety Commissioner, 2023).




There are risks associated with the various modalities that generative AI encompasses. Large language models which are text-based (e.g., chatbots) can be used to create highly convincing terrorist and violent extremist content or terrorist propaganda (Europol, 2023). Cases of perpetrators using generative AI to create child sexual abuse material and exploit children are also increasing (WeProtect Global Alliance, 2023).

The risk of impersonation (also known as 'deepfakes') increases when large language models are combined with other forms of generative AI, such as image or voice generators. Perpetrators can exploit the ability of large language models powered by AI to mimic natural human language. This capability allows offenders to groom children at-scale in automated and more targeted ways, with cases already reported where generative AI technologies are being used to facilitate child grooming (eSafety Commissioner, 2023). Annexure C – provides an overview of the rapid growth and development of AI over the last eight decades.

Open-source generative AI models whose code is freely available to all users present heightened risks, as users can modify the code to remove safeguards and tweak the model to enable the creation of harmful content such as child sexual abuse material (Clark, 2023). While there are real benefits to open-source innovation, research from the Stanford University's Institute for Human Centred AI has found clear risks of open-source generative AI foundation models being used to generate child sexual abuse material when compared to closed foundation models whose code is proprietary and not available to users (Kapoor, et al., 2024).

# 2. What are the objectives, why is the Government intervention needed to achieve them, and how will success be measured?

This chapter sets out the policy objective to be achieved; why Government intervention is needed to achieve the objective; the constraints and barriers to action; and how success will be measured.

## 2.1. The policy objective

eSafety's policy objective is to improve online safety for Australians in respect of class 1A and class 1B material – by ensuring that providers of RES and DIS services establish and implement systems, processes, and technologies to manage effectively the harms associated that Australians would solicit, generate, distribute, get access to or be exposed to class 1A material and class 1B material through their service.

This objective derives from sub-section 138(3) of the Act, which provides a list of examples of matters that may be dealt with by industry codes and standards and is consistent with the objectives of the registered industry codes (Codes Position Paper, 2021).

## 2.2. Why Government intervention is needed

Most class 1A and class 1B material depicts actions related to illegal activity or criminal offences. For example, the production, distribution, and possession of child abuse material are offences under certain Commonwealth and State and Territory legislation such as Division 273 of the Criminal Code Act 1995 (Cth) and section 51C and 51D of the *Crimes Act 1958* (Vic). However, the need for government intervention to manage the risks of Australians soliciting, generating, distributing, accessing, and being exposed to this material on RES and DIS operates as an additional safeguard to further strengthen this in practice, by imposing a set of mandatory compliance measures on certain providers.

While the RES and DIS sections are subject to a range of provisions in the Act, including schemes to require reporting against the Basic Online Safety Expectations (Part 4); cyber bullying material targeted at an Australian child (Part 5); non-consensual sharing of intimate images (Part 6); cyber abuse material targeted at an Australian adult (Part 7); material that depicts abhorrent violent conduct (Part 8); and the online content scheme (Part 9), these provisions do not address the risk of class 1A and class 1B material at a systemic level and the Act envisages that these schemes will be supplemented by either an industry code or industry standard for each identified section of the online industry.

The Basic Online Safety Expectations Determination 2022 (the Expectations) outlines the Australian Government's expectations that social media services and the RES and DIS sections will take reasonable steps to keep Australians safe. However, compliance with the Expectations is not mandatory. While eSafety can require providers to report on the steps they are taking to meet the Expectations, eSafety cannot compel compliance with the Expectations.

While eSafety is provided in Parts 5 – 9 of the Act with powers to direct the take-down of material from online services including RES and DIS under specified conditions, these schemes act retrospectively. The posting of the material must have happened (in addition to meeting other criteria) before eSafety can take any action. The schemes do not require RES and DIS services to proactively implement measures, systems or technologies which prevent the proliferation of class 1A and class 1B material (or indeed any other content).

For example, Part 9 of the Act gives the eSafety Commissioner power to give a removal notice to a RES or DIS in relation to class 1 material that is provided on its service. The notice can only be given after the material has been provided on the service, so it is necessarily after the fact. By requiring removal of material, these powers seek to address and reduce harm after the material has been shared online. However, the powers cannot prevent the posting of the material or impact its general availability. These powers therefore enable eSafety to alleviate the harm caused by the availability of such material online rather than prevent it surfacing in the first place. While other online industry sections are required by their industry code to take proactive measures to protect against the proliferation of the most harmful class 1A and class 1B material on their services, no such enforceable requirements exist for RES and DIS services.

As the alternatives to regulatory action – such as voluntary or co-regulatory schemes – have failed to deliver sufficient safeguards to meet the policy objective, Government intervention is therefore required.

### 2.2.1. Voluntary measures have failed to result in effective community safeguards

The widespread presence of child sexual abuse, pro-terror and violent extremist material on RES and DIS (as detailed in chapter 1) demonstrates that in the absence of regulation, industry participants will not voluntarily prevent the proliferation of this seriously harmful material on their services.

While some providers have taken steps to address online safety concerns on their service, these efforts vary between services and are often applied and enforced inconsistently across the multiple services offered by platforms.

Child sexual abuse material has continued to proliferate on many services despite the endorsement by many leading technology platforms of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse (Department of Home Affairs, 2020). These principles were developed by the Five Country Ministerial governments (Australia, Canada, New Zealand, the United Kingdom, and the United States) in consultation with a wide range of stakeholders including a leading group of industry representatives. This highlights the fact that public endorsement of voluntary principles does not mean that the companies will necessarily implement the effective policies and tools to achieve improved safety outcomes.

A recent OECD (2023) report examining the top 50 global online platforms' transparency reporting and policies and procedures in relation to child sexual exploitation and abuse found that 80 percent of platforms provided no detailed policy on online sexual exploitation of children and 60 percent of platforms did not issue a transparency report on such abuse.

OECD transparency reporting on terrorist and violent extremist content in the top 50 global online content sharing services also found that, while there has been an improvement in reporting and adoption of measures to prevent the upload and distribution of such material on several major content-sharing platforms following international calls for action from intergovernmental forums such as the Group of

Twenty (G20)[10], the Group of Seven (G7)[11] and the Christchurch Call[12], the measures are not adopted consistently across services (OECD, 2022).

On 29 August 2022, the eSafety Commissioner issued non-periodic reporting notices under section 56(2) of the Act to seven online service providers, requiring each provider to report on its implementation of the Basic Online Safety Expectations (the BOSE Expectations) with respect to child sexual exploitation and abuse. The information obtained in response to these notices provides valuable insights that have not been volunteered by providers, including in providers' own transparency reports (eSafety Commissioner, 2022). Further notices were issued in 2023 and 2024.

The 2022 notices were issued to the following providers responsible for the following services:

| Provider that received the section 56(2) notice | Services |
|---|---|
| **Apple Pty Ltd** | iCloud email<br>iCloud<br>iMessage<br>Facetime |
| **Meta Platforms, Inc** | Facebook<br>Messenger<br>Instagram |
| **WhatsApp LLC** | WhatsApp |
| **Microsoft Corporation** | OneDrive<br>Outlook.com<br>Xbox Live<br>Teams |
| **Skype Communications S.A.R.L** | Skype |
| **Omegle.com LLC** | Omegle |
| **Snap Inc.** | Snapchat |

[10] Members of the G20 are Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Indonesia, Italy, Japan, Republic of Korea, Mexico, Russia, Saudi Arabia, South Africa, Turkiye, the UK, the US, the African Union and the European Union

[11] Members of the G7 are Canada, France, Germany, Italy, Japan, the UK and the US

[12] A reference to the Christchurch Call to Action Summit initiated by New Zealand and held on 15 May 2019 in Paris two months after the Christchurch Mosque shootings, at which a pledge was signed by 54 governments and 8 online service providers as part of the Global Internet Forum to Counter Terrorism (GIFCT)

Services offered by these providers will be covered by the standards with messaging, chat and gaming services covered by the RES standard and photo and storage services (as well as other websites and apps) covered by the DIS Standard.

Providers were asked specific questions about the tools, policies and processes they are using to address various forms of child sexual exploitation and abuse, such as the proliferation of this material online, the online grooming of children, and the use of video calling and conferencing services to provide live feeds of child abuse.

As a result of the information provided in response to the notices, eSafety found:

> ‘Significant variation in the steps being taken by providers to protect users and the wider Australian public. There is no common baseline, either between providers or even across a provider’s own services. For example, while eSafety found some providers use well established ‘digital fingerprinting’ technology tools to identify images or videos previously identified as being CSEA material across all the services eSafety asked about, other providers use these tools on some of their services, but not others. These tools have an error rate of about 1 in 50 billion. Until now, providers have not been open about these differences.
>
> Some providers are checking for new or ‘unseen’ CSEA [child sexual exploitation and abuse] material, or using technology to detect potential grooming conversations, while eSafety was told by another provider that there is no technology good enough for either purpose. Most providers who were asked did not identify specific steps being taken to identify the abuse of children through live video calls, conferences, or streams.
>
> There is significant variation in the steps being taken to prevent recidivism (where users banned for previous abuse re-register with new accounts). Some providers report tracking extensive lists of indicators of recidivism, while others report only using a minimal number. There is also significant variation about what information is shared between a provider’s own services to prevent banned users operating on multiple parts of a provider’s products.

> There are significant differences in the speed with which providers respond to user reports of child sexual exploitation, with responses varying from 4 minutes to 2 days (and 19 days where eSafety were told cases needed 're-review'). Some other providers have no reporting options at all within the app or service, requiring users to contact the provider via e-mail if they wish to complain about illegal or harmful activity on a service.'
>
> (eSafety Commissioner, 2022)

As previously detailed in chapter 1, online harms continue to increase highlighting that self-regulation is not an effective means of combatting the proliferation of harmful class 1A and class 1B material on RES and DIS.

### 2.2.2. Attempted co-regulation failed to result in appropriate community safeguards

Industry-developed draft codes for RES and DIS were not registered by the eSafety Commissioner because they did not contain appropriate community safeguards, a statutory requirement for registration under paragraph 145(1)(a)(ii) of the Act (eSafety Commissioner, 2023).

The draft RES industry code did not provide appropriate community safeguards because:

- there was no requirement on closed communication and encrypted RES Providers with the sufficient capability to detect and remove known (i.e., pre-identified) child sexual abuse material and known pro-terror material;
- the requirements on certain RES Providers to act and invest in disruption and deterrence of child sexual abuse material and pro-terror material failed to address the differing capabilities and functionalities of RES resulting in a very low bar for compliance for many RES Providers.
- there was no requirement on closed communication RES Providers (such as email providers) to have trust and safety personnel;
- there was no requirement on certain RES Providers (those which consider themselves to be not capable of reviewing and assessing materials on their services) to enforce their own policies relating to class 1A and 1B material.

The draft DIS industry code did not provide appropriate community safeguards because:

- there was no requirement on end-user managed hosting services to:
    - deploy systems, processes and/or technologies to detect and remove known (pre-identified) child sexual abuse material and known (pre-identified) pro-terror material;
    - act and invest in disruption and deterrence of class 1A material (including new/first generation child sexual abuse material).
- there was no requirement for certain end-user managed hosting services (those which consider themselves to be not capable of reviewing and assessing materials on their services) to enforce their own policies or terms of use relating to class 1A and 1B material;
- it did not adequately address measures directed towards achieving the objective of ensuring that industry participants have scalable and effective policies, procedures, systems, and technologies in place to take reasonable and proactive steps to limit the hosting of class 1A material and class 1B material in Australia.

If RES and DIS industry sections are not subject to enforceable requirements to address the risk of such material on their services, there is a significant risk of harm due to the rapid proliferation of material containing illegal and restricted material on RES and DIS. The failure of the industry draft codes to meet requirements leaves a gap in Australia's regulatory framework, which as the Act envisaged, requires intervention by the eSafety Commissioner in the form of industry standards.

### 2.2.3. Governments globally recognise the need for intervention

Since the Act came into effect in Australia, multiple overseas governments have also concluded that voluntary regulation by industry has failed to adequately protect their citizens from the proliferation of high-risk online harms, and have either introduced, or are in the process of introducing, legislation to regulate the online industry.

The United Kingdom's Online Safety Act 2022 (UK OSA) (Online Safety Act 2023) which came into effect in October 2023 sets out key online safety measures that

align with Australia's approach. Like the RES and DIS standards, the UK OSA includes obligations on services to prevent and remove illegal content, with requirements to report identified child sexual exploitation material to law enforcement agencies and/or verified organisations and a requirement to conduct a risk assessment before making any significant or material changes to the service. Reforms announced on 16 April 2024[13] will also criminalise the creation of sexually explicit 'deepfake' images of adults without consent through an amendment to the Criminal Justice Bill with an unlimited fine.

The European Union's *Digital Services Act* (European Commission, 2024) commenced in February 2024. Although it has a much broader scope than our Online Safety Act and a different framework, it does include some similar measures to those contained in the standards and codes in terms of empowering end-users and increase the responsibility of service providers. The Digital Services Act requires service providers covered by the Act to remove illegal content, have easily accessible and clear content reporting mechanisms to enable end-users to report illegal content, and to publish in plain language their service terms and conditions.

Ireland's Online Safety Media and Regulation Act 2022 includes provisions to address the regulation of content for online safety. Following the commencement of the first Online Safety Commissioner, online safety codes have been proposed to address child sexual abuse material and terrorist material on social media services, with the regulator having powers to assess the compliance of online services with the safety codes.

In February 2024 the Canadian Government introduced Bill C-63 in Parliament to create a Canadian *Online Harms Act* requiring mandatory reporting of 'internet child pornography' by service providers. If approved, the Canadian Act will provide a baseline standard for online platforms to keep Canadians safe by holding online platforms accountable for the content they host. Bill C-63 proposes stronger protections for children online and better safeguards for Canadians from online hate. It specifically targets several types of harmful content: including content that sexually victimises a child or revictimizes a survivor; content that incites violence; and content that incites violent extremism or terrorism.

---

[13] See https://www.gov.uk/government/news/government-cracks-down-on-deepfakes-creation

Given the global reach and operations of large online participants, international cooperation, and collaboration on online safety issues by governments and regulators is critical. The Global Online Safety Regulators Network (the Network) has been established to bring together independent regulators from across the world to cooperate across jurisdictions and to share information, best practice, experience, and expertise, and to support harmonised or coordinated approaches to online safety issues. (Global Online Safety Regulators Network, 2022).

Specific guardrails are also being put in place to address risks and harms associated with generative AI technology (which go further than those proposed in the standards).

## 2.3. Constraints and barriers to achieving the objective

The introduction of regulatory requirements must be undertaken with a clear awareness of constraints and barriers, both actual and potential. There are several significant constraints and barriers to improving online safety for Australian end-users in respect of class 1A and class 1B material on RES and DIS.

### 2.3.1. The scale and global nature of the problem

The global nature of the internet and the significant number of providers based overseas with online services accessible in Australia creates challenges for compliance and enforcement of the Act as a while. It is also a reason why a 'whole of stack' approach was taken to Part 9 of the Act (which requires industry codes or standards across the online eco-system) and why eSafety's takedown schemes need be accompanied by ex-ante regulation requiring proactive steps by each industry section to address systemic issues.

The large number of online services, and the wide variety of services, within the RES and DIS sections, make regulation and enforcement difficult, and places a consequential administrative impost on the relatively small Australian regulator which has finite resources. To address this and to ensure a proportionate approach to risk, the reporting requirements in the standards (regarding risk assessments, technical feasibility, development program outcomes and annual

compliance reports) only mandate reports in a small number of cases, in other cases reports can be required on request by eSafety.

Separately, the investment obligations in the standards require only those services with a minimum number of monthly active users to have an investment and development program in place to disrupt and deter child sexual abuse and pro-terror material. The monthly active user threshold was given careful consideration by eSafety and we believe it is appropriate and proportionate to not burden smaller providers. The threshold does however create a limitation as those that fall on the outside of the threshold will not have to comply with the investment obligation.

### 2.3.2. Regulation needs to keep up with technological innovation

The rapidly evolving nature of the online environment is a key challenge for regulation. The constant development of new technologies and the introduction of new functionalities and features creates challenges to compliance and enforcement challenges. For example, the rapid evolution of generative artificial intelligence has introduced new risks given the new opportunities to create class 1A and class 1B material and, as addressed in response to question 1 above. However, as eSafety has previously acknowledge AI can also be harnessed to significantly improve current proactive content moderation technologies to quickly and accurately address harmful material (eSafety Commissioner, 2023).

In Australia, the Government is looking at the risks, benefits, and potential impacts of generative AI. On 17 January 2024 the Department of Industry, Science, and Resources (DISR) published its interim response to the safe and responsible AI consultation held in 2023. Feedback on the interim response is to inform consideration across government on appropriate regulatory and policy responses. Targeted joint work has also been carried out by the Digital Platform Regulators Forum (DP-REG), which includes the Australian Competition and Consumer Commission (ACCC), Australian Communications and Media Authority (ACMA), Office of the Australian Information Commissioner (OAIC), and eSafety[14].

[14] Digital Platform Regulators Forum, *Working Paper 2: Examination of technology – Large Language Models* https://dp-reg.gov.au/publications/working-paper-2-examination-technology-large-language-models

In view of the rapidly evolving technical landscape, the standards take a technology-neutral approach to implementation, identifying outcomes rather than prescribing the technology to be used, and ensuring there are proportionate obligations across technology ecosystems. The DIS Standard focusses on key risk areas in relation to generative AI services.

### 2.3.3. Perpetrators' obfuscation and evasion techniques

The standards contain a suite of complementary obligations that ensure a robust and effective approach to address the systemic issue of Class 1A and 1B material on these services. This is necessary given it is not possible for one measure to address the future tactics of malicious actors (e.g. those creating, sharing and storing child sexual abuse material). Further, as set out above, the standards require some services to establish and implement development programs and invest in systems, processes, and technologies to enhance the ability of service providers to detect and disrupt child sexual abuse material and pro-terror material online. This is important because as perpetrators develop tactics in response to existing safeguards (WeProtect Global Alliance, 2023), services must continue to invest in the safety of their services.

## 2.4. How will success be measured?

The objectives of the standards are to improve online safety for Australians in respect of class 1A material and class 1B material by ensuring that providers of relevant electronic services establish and implement systems, processes and technologies to manage effectively risks that Australians will solicit, generate, distribute, get access to or be exposed to class 1A material or class 1B material through the services.

Critical to this is the risk based, proportionate approach to the requirements in the standards, the complementary suite of measures in each standard and the enforceability of the requirements. Services with substantial reach used by many Australians every day will be covered by these standards and will be required to take proactive steps to address these harms. eSafety will focus on encouraging compliance by providers and across the eco-system at large to combat systems risks associated with class 1A and 1B material. Success will therefore be achieved through RES and DIS providers engaging with the standards, improving their safety

practices, and proactively addressing systemic issues to reduce the risk of class 1A and class 1B material on their services. The measures of success will include:

- **RES and DIS providers engaging with the standards.** RES and DIS providers who send eSafety timely annual reports and risk assessments (as required by the standards); who are responsive to notices issued by eSafety; who proactively notify eSafety of new features or functions which may present an increased safety risk in respect of class 1A or class 1B material; and who are responsive to informal requests[15] from eSafety for the removal of class 1A and class 1B material – are demonstrating through these behaviours their underlying commitment to the policy objectives.

- **Certain known class 1A material is proactively detected and removed by RES and DIS providers.** There is currently no industry baseline for the proactive detection and removal by RES and DIS providers of known child sexual abuse material and pro-terror material on their services. The new requirements for proactive detection and removal in the standards are expected to increase the deployment of technology and systems to proactively detect and remove the material in forward years. This will be ascertained by compliance activities and the annual compliance reports submitted.

- **Positive safety interventions have been taken by RES and DIS providers.** Across the reporting period eSafety will track the introduction of online safety interventions by RES and DIS providers which can be wholly or partially attributed to the standards, such as introduction of user reporting options, through reports provided and such periodic BOSE notices as may be issued.

- **Feedback from stakeholders on the effectiveness of the RES and DIS industry standards.** Feedback from stakeholders as to whether they consider the standards are effective in increasing online safety in respect of class 1A and class 1B material across RES and DIS services. Stakeholders could include (but are not limited to) the National Centre for Missing and Exploited Children, Tech Against Terrorism, researchers, academics, and community safety advocates.

[15] An 'informal request' refers to a request which is made without issuing a formal notice under the Act or the relevant standard. Compliance by a provider with an informal request without the need for eSafety to issue a formal notice which may attract a penalty if not complied with is a sign of the provider's engagement with the standards and commitment to the underlying safety objectives of the regulatory framework.

The evaluation metrics for these success measures can be seen in Table A in chapter 7 of this document.

# 3. What policy options are being considered?

**This chapter examines three options to achieve the policy objective, provides an overview of each option and explains how it was developed. The three options are:**

- **Option 1 – maintain the status quo.**
- **Option 2 – co-regulation; and**
- **Option 3 – Government intervention.**

## 3.1. Option 1 - maintain the status quo

Option 1 (maintain the status quo) represents the baseline or no-change option. Option 1 would see no additional regulation – either through an industry code or through a standard - regarding the treatment of class 1 material for RES and DIS. As set out above, codes containing appropriate safeguards in respect of class 1A and class 1B material have been registered under Part 9 of the Act and commenced for social media services; internet carriage services; equipment providers; app distribution services; hosting services; and internet search engine services.

Option 1 would mean RES and DIS would have a lower level of regulation than the six other industry sections for which codes have been registered by the eSafety Commissioner. RES and DIS would be the only online industry sections where there is no legal requirement to proactively address the serious harm caused to Australians by the generation, hosting, and distribution of the most harmful online material.

Under option 1, the eSafety Commissioner would have access only to the statutory powers currently available in respect of class 1 content on RES and DIS. These powers are:

i. **Removal of specified class 1 material under the Online Content Scheme in Part 9 of the Act** –by giving a removal notice under section 109 requiring a RES or DIS to take all reasonable steps to ensure the removal of specified class 1 material from their service within 24 hours (or such longer period as the Commissioner allows) or face a civil penalty of 500 penalty units (section 111). The Commissioner can issue a formal warning if the service fails to pay the penalty (section 112).

ii. **App removal** – under section 128 of the Act, if an app distribution service (app store) which enables end-users in Australia to download an app that facilitates distribution of class 1 material the eSafety Commissioner may give an app removal notice requiring the app distributor to, within 24 hours (or such longer time permitted by the Commissioner), cease enabling end-users in Australia to download the app, or face a civil penalty of 500 penalty units (section 129). An app removal notice may only be given where the Commissioner is satisfied there were 2 or more times during the previous 12 months when end-users in Australia could use the service to download the app, and during the previous 12 months the Commissioner issued one or more removal notices under section 109 for class 1 material distribution facilitated by the app which were not complied with. As such, the app removal notice requires evidence of a certain degree of ongoing harm to issue a notice rather than a single instance.

iii. **Link deletion** – under section 124 of the Act, if class 1 material is accessible via a link on a search engine service, the eSafety Commissioner may issue a link deletion notice to the search engine requiring the search engine to, within 24 hours (or such longer time as permitted by the Commissioner), cease providing a link to the service, or face a civil penalty of 500 penalty units (section 125). A link deletion notice may only be issued where the Commissioner is satisfied that there were 2 or more times during the previous 12 months when end users could access class 1 material using a link provided by the service and during the previous 12 months, the Commissioner gave one or more removal notices under section 109 or 110 in relation to class 1 material that could be accessed using a link provided by the service that were not complied with. Like the app removal notice, a link deletion notice requires evidence of systemic harm rather than single instance before it can be issued.

iv. **Service provider notifications** – under section 113A of the Act the eSafety Commissioner can publish a statement on the eSafety website where a RES or DIS service has on 2 or more occasions during the previous 12 months had class 1 material on its service which contravened the service's terms of use and give a copy of the statement to the service provider.

v. **Application for an order to cease** – the eSafety Commissioner may apply to the Federal Court for an order to require a provider to cease providing a RES (section 157) or a DIS (section 158) where the Commissioner is satisfied that the RES or DIS on 2 or more occasions during the previous 12 months contravened a civil penalty provision under Part 9 of the Act and as a result the continued operation of the RES or DIS represents a significant community safety risk. Whilst this approach is available to the Commissioner it is subject to stringent statutory thresholds, meaning this power would be reserved for certain circumstances.

vi. **The Basic Online Safety Expectations (BOSE) scheme would continue to apply to the RES and DIS sections** - Under the Act, the eSafety Commissioner can require reporting on how a provider is meeting any or all the Expectations. While the obligation to respond to a reporting requirement is enforceable and backed by civil penalties, the Expectations are themselves are not mandatory, unlike industry codes and industry standards.

Option 1 would mean that eSafety's ability to act in respect of class 1A and class 1B material on RES and DIS would be limited to only material which has been notified about or become aware of under one of the existing schemes in the Act (outlined in i-vi in the preceding paragraph).

Option 1 does not place any requirements on RES and DIS providers to proactively address class 1A and class 1B material. RES and DIS providers therefore would have no enforceable obligations to:

- detect and remove, or disrupt and deter some class 1A material (child sexual exploitation and pro-terror material)
- ensure systems and processes are in place to respond to terms of service breaches regarding class 1A and class 1B material
- incorporate safety features and settings that minimise the risk of class 1A and class 1B material on their service
- maintain sufficient trust and safety functions and personnel

- provide a complaints mechanism for end-users and account holders to report class 1A and class 1B material
- carry out risk assessments to determine the risk of class 1A and class 1B material on the service, and
- proactively provide regular reports to eSafety on key safety issues.

Option 1 would mean that the increasing availability of harmful child sexual abuse material, pro-terror material and other class 1A and class 1B material on RES and DIS could not be managed at scale by the regulator (eSafety).

Option 1 does not meet the policy objective, as it fails to provide appropriate safeguards in respect of the creation, hosting, sharing and proliferation of the most dangerous and harmful online material via RES and DIS.

## 3.2. Option 2 - industry co-regulation

Option 2 (industry co-regulation) would require the registration of RES and DIS industry codes for class 1A and 1B online material which are able to be registered by the eSafety Commissioner. While draft industry codes have been developed, these were not registered as they did not provide appropriate community safeguards. This was a legislative requirement because the matters that the draft code dealt with were all matters that the Commissioner considered to be of substantial relevance to the community.

Part 9, Division 7 of the Act allows for the establishment of new industry codes or standards to regulate sections of the online industry. The Act provides for industry bodies or associations to develop, and eSafety to register, the new industry codes.

In September 2021 eSafety issued a Position Paper to assist industry associations to prepare draft codes. The Position Paper drew on eSafety's engagement with industry and was informed by a review of local and international regulatory approaches, engagement with industry bodies and associations, and with national regulators with interconnected regulatory schemes. It outlined the expectations for the development by industry associations of codes, as well as eSafety's preferred outcomes-based model for the codes[16].

[16] at https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf

The Position Paper was clear that, 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 codes that contain adequate mechanisms for preventing or limiting online material containing illegal and restricted 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.

Part 9 of the OSA provides that if appropriate codes cannot be established, the eSafety Commissioner has the power under the Act to declare standards.

The RES and DIS draft industry codes were developed by industry associations representing RES and DIS providers over a period of two years. During this period, eSafety provided considerable feedback and engaged extensively with industry.

The eSafety Commissioner formally declined to register the draft RES and DIS industry codes in May 2023 on the basis that they did not provide appropriate community safeguards in relation to matters that they dealt with. An overview of the statement of reasons for this decision is provided above in section 2.2.2. and the full statement of reasons for the decision to refuse to register the draft industry developed codes is available on the eSafety website[17].

Further discussion and development by industry is not expected to result in RES and DIS codes which meet the statutory requirements and delays putting in place effective requirements to protect the community in respect of class 1A and class 1B material.

## 3.3. Option 3 - direct regulation

Option 3 is that the eSafety Commissioner register the standards, putting in place obligations on services covered by these standards.

The eSafety Commissioner is empowered under section 145 of the Act to determine industry standards through the creation of a legislative instrument if the industry developed draft code does not contain appropriate community safeguards, or in other circumstances detailed in the Act. As set out above, the

[17] at https://www.esafety.gov.au/industry/codes

eSafety Commissioner formally declined to register the draft RES and DIS codes on 31 May 2023 on the basis that they did not provide appropriate community safeguards in relation to the matters they dealt with.

Regulation through industry standards will provide adequate regulation of providers of RES and RES to reduce the risk of class 1A and class 1B material on their service. Regulation via registration of the DIS Standard and the RES Standard is consistent with the objectives of the Act in section 3 and the eSafety Commissioner's statutory functions in section 27 of the Act.

The standards are required to ensure that the RES and DIS sections of industry provide a similar level of protection as some of the other online industry sections which have registered codes in place.

As set out above, the standards will operate alongside the six registered industry codes and impose a set of mandatory compliance measures, requiring service providers to:

- take proactive steps to create and maintain a safe online environment;
- empower end-users in Australia to manage access and exposure to class 1A and class 1B material; and
- strengthen transparency of, and accountability for, class 1A and class 1B material on their services.

As set out above and consistent with the already registered industry codes, the draft standards adopt an outcomes- and risk-based approach. The requirements in the standards are proportionate to the risk a service presents in respect of class 1A and 1B material. The requirements are also outcomes-based, in that they set out what they are intended to achieve while providing flexibility in how those outcomes are to be achieved. This approach recognises that:

- different services and technologies may have different risk profiles;
- compliance measures should be proportionate to the level of risk associated with a particular service which considers a range of factors including the reach of the service; and
- compliance measures should be flexible, to enable effective implementation, recognising the differences between unique services, and to adapt to changes in technology and in the risk environment.

In developing the standards, eSafety built on the extensive work of industry bodies in developing and consulting on the draft RES Code and DIS Code. eSafety used provisions of the draft codes as an initial base for standards requirements, while addressing the deficiencies identified also developing new measures to address risks posed by generative AI.[18] Submissions received on the draft standards from industry, civil society groups, government agencies and other interested parties in December 2023/January 2024 have also been closely considered by eSafety. Multiple changes were made to the final standards in response to this feedback.

Option 3 will put in place new regulatory requirements on certain RES and DIS including:

- Requirements to conduct risk assessments to reduce the risks of class 1A and class 1B material being generated, posted, stored, or distributed on RES and DIS services. The standards require providers of certain services to self-assess their risk to identify their risk tier and consequent legal obligations.
- Specific obligations on certain service providers in relation to 'known' child sexual abuse material and pro-terror material (that is images and videos that have been verified as such)[19] and 'new' cases of such material, including:
    - requirements for certain RES and DIS providers to proactively detect and remove known child sexual abuse material and known pro-terror material, where identified limitations do not apply;
    - requirements to take appropriate alternative action where it is not technically feasible or reasonably practicable to deploy tools to automatically detect and remove known child sexual abuse material and known pro-terror material on the service, and
    - obligations for certain RES and DIS providers to take action to disrupt and deter end-users from using the service to solicit, create, post, or disseminate both new and known child sexual abuse material and pro-terror material.

[18] For background information on generative AI and the online safety risks associated with this technology, see eSafety's Tech Trends position statement on generative AI.
[19] See footnote 3 for definition.

- Requirements on:
    - Pre-assessed and Tier 1 RES services with more than 1 million monthly active users in Australia;
    - Tier 1 DIS services and high impact generative AI DIS with more than 1 million monthly active users in Australia; and
    - end-user managed DIS hosting services with more than 500,000 monthly active users in Australia;
- to have a development program including investment in respect of systems, processes, and technologies to detect and identify and disrupt and deter child sexual abuse material and pro-terror material on the service.
- Requirements for certain RES, including communication RES, to take appropriate action to engage with reports of class 1A and 1B materials and determine whether terms of use or policies have potentially been breached.
- Specific obligations on model distribution platforms, and specific obligations on high impact generative AI DIS providers where there is a material risk that end-users can generate material which would be classified as X18+ or RC.[20]

The standards involve a suite of targeted requirements which allow them to adapt to emerging technologies, services, and operating practices while still ensuring regulatory measures are proportionate and appropriate to the level of risk a service poses. This approach will provide flexibility in the face of new variations in online harms as well as the emergence of new safety technologies and best practises.

The regulatory approach underpinning option 3, adoption of the standards, is:

- **risk-based** – the obligations in the DIS Standard and RES Standard are tailored and focused on those services where the greatest risk of harm arises. Those providers which do not fall within a pre-assessed or defined category will also be required to conduct a risk assessment to determine the risk profile of their service(s). The risk is to be assessed by factors which predict the likelihood of harm a service poses to the end-user and

[20] These obligations are broadly consistent with emerging generative AI best practise, including with the industry back report from Thorn and All Tech is Human titled 'Safety by Design for Generative AI: Preventing Child Sexual Abuse' (2024), as well as with the National Institute of Standards and Technology's 'Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile' (April 2024).

the potential severity of this harm. This allows the RES and DIS standards to target risks appropriately and be proportional in mitigating them. A risk-based approach is beneficial to improve compliance outcomes, as it tailors each service provider's obligations to their level of risk allowing services to focus on their specific requirements at a reduced regulatory cost burden. (NSW Finance, Services and Innovation, 2016) Regulation that does not effectively target the causes of risks, often fails to deliver any real benefits and results in higher cost burdens for providers. (OECD, 2021) The standards therefore require those services with a higher likelihood of harm to comply with more stringent obligations that lower-risk services are not subject to;

- **proportionate to the assessed risk of the service** – to reduce any unnecessary compliance burden and ensure obligations are appropriately attached to the level of risk of class 1A or class 1B material on a service. Given the vast scale of the internet and the large number of service providers that fall under the standards, a risk-based approach is best suited to regulation of online service providers as it allows for flexibility and is context responsive given the significant spectrum of risk profiles; and
- **outcomes and principles-based** – the standards do not rely on prescriptive rules, instead focussing on the outcomes that must be achieved to decrease harms for Australian end-users on RES and DIS services. This encourages innovation as companies are required to develop solutions and create their own processes and mechanisms to comply with the outcomes. This places the onus on companies to create meaningful solutions, rather than simply meeting the basic requirements. This flexibility empowers them to make their own choices around the specific systems, processes, and technologies they implement that add value to their service as well as comply with the standards.

This approach lowers the regulatory burden by not requiring a one-size-fits-all approach which allows services to best tailor their approach and investment to suit their individual needs. As technology rapidly changes in these dynamic digital industries, this outcomes and principles-based approach is designed to drive continuous improvement and best practice.

# 4. What is the likely net benefit of each option?

**This chapter estimates the likely net benefits of the options being considered. The regulatory cost burden is estimated for the relevant population, using benefit transfer methodology on secondary source data in place of independent cost benefit analysis to derive estimates.**

## 4.1. Methodology

Estimating the regulatory burden costs for RES and DIS providers in scope of this analysis is impacted by several factors (see section 4.1.2.3).

The regulatory burden costs were established through a benefit transfer methodology using secondary source data in place of independent cost-benefit analysis. The benefit transfer methodology enables the use of data from already completed studies in other locations and/or contexts (in this case from the United Kingdom's 2022 assessment of the impacts of its Online Safety Bill) (herein referred to as the UK OSA) to estimate economic values or other costs.

Benefit transfer is a methodology that can be used when it is too expensive and/or there is too little time available to conduct an original valuation study, yet some measure of benefits is needed to be determined. As no specific cost estimates were provided by industry participants with the draft industry codes[21] this impacted establishing the provision of cost estimates specific to the complex range of RES and DIS services with obligations under the proposed policy and options.

The key limitation of this methodology is that that benefit transfers can only be as accurate as the initial study and values may not be comparable on all measures. This UK OSA is comprehensive piece of legislation that extends much further than the policy options in this impact analysis.

This methodology was adopted to estimate the economic and regulatory burden costs to key stakeholder groups due to the absence of available data and limited

[21] Although eSafety received feedback from some providers that compliance with the standards would be financially onerous during industry consultation, no actual estimates were provided to eSafety despite being requested.

resources to undertake independent cost-benefit analysis. Estimates are calculated on a best-efforts basis.

### 4.1.1. Methodology for estimating the number of businesses in scope

There is no existing data on the number of services within scope of the policy options. Due to the wide range of services captured, and the lack of data available on RES and DIS providers, it is difficult to estimate the number of online services likely to be affected by the policy options. For example – services considered to be RES range from messaging applications, gaming platforms, dating services, telephony RES (SMS/MMS) and enterprise services. DIS encompasses any website or app that is not a RES or social media service, ranging from generative AI services to pornography sites, and cloud storage services.

Many providers with obligations under the policy options are international businesses that operate in and/or provide RES and DIS services that are accessible to and used by end users in Australia. Large-scale international operators may also provide multiple services (i.e., social media and messaging) and have a higher number of employees (e.g., > 50,000) compared to even the largest Australian RES and DIS with obligations under the Standard.

The impact on international (i.e. overseas based) RES and DIS - although covered by the policy options - are not included for the purposes of this regulatory burden estimates. This is to align with the methodology used in the secondary source data, which considers the costs to UK businesses[22]. More information on methodology, data sources and assumptions underpinning the estimates for businesses in scope are provided in Annexure B.

#### 4.1.1.1. The total estimated number of impacted businesses in scope

Data was collected from the Australian and New Zealand Standard Industrial Classification (ANZSIC) codes, Australian Bureau of Statistics (ABS) reporting, and other government and open-source data collections to capture RES and DIS subject to obligations and likely to incur regulatory costs under the policy options. Due to the variability in services captured under the policy options it was

[22] As per guidelines from the HM Treasury Green Book, the UK impact assessment only considers effects on UK businesses.

determined that a range of sources was required to provide the most accurate representation and capture of impacted Australian RES and DIS service providers. Where possible data was sourced from Australian government agency reporting and supplemented with other open-source data. A breakdown of these sources is provided in Annexure B.

Combining these data sources estimate that there would be approximately 6.7 million Australian RES and DIS providers potentially in scope at the end of the 10-year appraisal period. Most of this figure encompasses Australian websites (n=6.6 million) which the majority of will not have any meaningful obligations/regulatory burden costs under the policy options and have been deducted from the regulatory burden estimates (these businesses are likely to be Tier 3 under the DIS Standard).

A proportion of websites was included as a 'representative estimate' of Australian DIS that may fall have obligations – for example high-impact Australian based services hosting X18+ or R18+ content, file/image storage services or high-risk generative AI services.[23]

Based on these data sources it is estimated that there are **2,045** Australian RES and DIS likely to incur regulatory costs at the end of the 10-year period (including compound growth measurement). This baseline has been used to estimate the regulatory burden costs under section 4.1.2.

A breakdown of assumptions underpinning the baseline services in scope is provided in Annexure B.

## 4.1.2. Methodology for estimating regulatory burden costs

### 4.1.2.1. Relevant population for assessing costs

In accordance with Government's Regulatory Burden Measurement framework[24], and the scope and parameters of the governing regulatory framework (the Act and its associated codes and standards), the relevant population for the purposes of quantifying costs is as outlined in Figure 1 below.

[23] Most Australian porn sites are assessed to pose a much lower risk for CSAM than international porn sites because they are run by small businesses/independent operators who produce all the content (there is no user generated content) and are pay-to-access.
[24] see https://oia.pmc.gov.au/resources

**Figure 1 – Relevant population for assessing costs**

| Stakeholder | Definition |
|---|---|
| Individuals | A person subject to Australian law, whose activities have an impact in Australia and who is affected by the proposed policy, and who accesses or may access RES or DIS in Australia. |
| Community organisation | Any organisation engaged in charitable or other community-based activity operating under Australian law and not established for the purpose of making profit. |
| Businesses | Australian RES and DIS providers |

#### 4.1.2.2. Calculation of the regulatory cost burden

Drawing on the impact assessment of the comprehensive UK OSA[25] , which was completed in January 2022, compliance elements were identified that were transferable to these policy options (ie compliance elements that related to the class 1A and class 1B material risk mitigation measures in the standards or draft codes) and to some of the types of service providers in scope of policy options. The UK assessment considered comparable timeframes, broadly similar demographics, and levels of technological infrastructure. Capital and labour costs between the UK and Australia are also comparable.[26]

There are however substantial differences in the scope of the UK legislation including a significantly wider range of harms and obligations in the UK than for options 2 and 3, service-types, and the nature and scope of compliance obligations on different service-types which impact the regulatory cost. For example, the UK OSA costings include obligations in relation to actioning a wider range of material and harms (such as fraudulent advertising, children's access to online pornography, cyberbullying, image-based abuse, cyberstalking, and protection of content of democratic importance) whereas the policy options only address class 1A and class 1B material). Although these differences impact the accuracy of any benefit transfer to the Australian environment, for the purpose of a best-efforts estimation they can still be used to provide an indication of the

[25] As a member of the Organisation for Economic Co-operation and Development (OECD) the UK follows a robust regulatory framework, requiring impact assessment to inform government decision making processes where government intervention/regulation is required.

[26] Extensive research was undertaken by the UK Department of Digital, Culture, Media and Sport, including the engagement of external consultants, rapid evidence assessments, business engagement and evidence (costs) requested from industries and businesses in scope of regulatory burden. The costs and benefits provided by the UK impact analysis are illustrative and intended to provide an indication of the likely scale of impact from primary and secondary legislation and future codes of practices. An overview of the collection and methodology used to assess their impact evaluation is further provided in Annexure B.

likely scale of impact from the introduction of the policy options. A key limitation of this method is also that it does not canvass and cost the totality of the compliance obligations under the policy options – only those which are reflected in the UK assessment. Nevertheless, it is considered to cover the key obligations and obligations that would have the most significant regulatory burden (i.e. deploying proactive content moderation technologies).

In accordance with Government's Regulatory Burden Measurement framework regulatory burden costs are presented as average annual impacts and costed over a 10-year default duration of the policy (compound growth calculated). As varying costs are expected, the average annual impact is calculated by dividing the total estimated cost over 10 years by this timeframe. Costs are presented in real terms (also referred to as constant prices) as average annual figures and not adjusted for inflation within the 10-year period[27] Also in accordance with the Government's Regulatory Burden Measurement framework, while compliance costs are estimated, enforcement costs are excluded.

The steps to calculate the estimate of costs for the policy options (costed options 2 and 3) were as follows:

- The compliance obligations outlined in the UK OSA impact assessment[28] over a 10-year appraisal period (starting from the date of the UK OSA implementation) were mapped against the most closely comparable obligations for option 2 and 3.
- The costs provided in the UK OSA impact assessment were adjusted for inflation and Australian exchange rates.[29]
- The total costs of the comparable UK obligations over the 10-year appraisal period were added to arrive at an overall comparable cost estimate of compliance option 2 and 3 over a 10-year period, based on the estimated number of RES and DIS in scope.
- The costs for option 2 and 3 were then further scaled as a proportion of the costs of the UK OSA impact assessment. The total costs were adjusted to reflect a broad qualitative assessment of the enforceability and scope of

[27] Inflation has been applied to the 2019 UK costs to bring them to 2023 AUS costs figure. Inflation was calculated used the Reserve Bank of Australia tools.
[28] UK Online Safety Bill Impact Assessment https://assets.publishing.service.gov.uk/media/6231dc9be90e070ed8233a60/Online_Safety_Bill_impact_assessment.pdf
[29] Inflation rates calculated using the RBA inflation calculator, exchange rates dated 15 March 2024

each policy option as proportion of the UK OSA obligations and costs (refer to Annexure B for more detail).

- All the estimates provided (total, annual and per-business) is based on total costs at the end of a 10-year period.

#### 4.1.2.3. Key factors impacting regulatory burden estimates.

It is expected that costs will differentially impact those RES and DIS with obligations under Option 2 and 3 based on the service(s) provided, the risks of class 1 material (higher risks require more meaningful obligations) and may in some cases be disproportionate to the size/revenue of the business. Therefore, the following factors should be considered alongside the regulatory burden estimates in section 4.2:

##### i. There is no baseline for business-as-usual costs.

Under the Government guidelines, business as usual activities are excluded from regulatory burden costing. However, it is difficult to exclude these costs accurately as they are unknown. There is no available data or research in Australia or internationally which quantifies the level of existing mitigations that RES and DIS providers already have in place to manage class 1A and 1B material on their services. Transparency notices issued by eSafety in 2022 and 2023 to providers that offer RES and DIS confirmed that some global RES and DIS already have systems, processes, and technologies in place (eSafety Commissioner, 2022). However, these have been implemented incompletely and inconsistently across services. Table 1 below provides a summary of some of the results received from the transparency notices in 2022 and 2023 on international RES and DIS who offer services to Australian end-users.

**Table 1 – Results from 2022 and 2023 BOSE notices - RES and DIS and the mitigations in place for the detection and removal of child sexual abuse material (NB: CSEA refers to child sexual exploitation and abuse).**

| Company | RES or DIS | Uses hash matching to detect known CSEA images | Uses hash matching to detect known CSEA video | Uses tools to identify new CSEA images |
|---|---|---|---|---|
| Apple | iCloud | No | No | No |
| Apple | iCloud email | Yes | No | No |
| Apple | iMessage (E2EE by default) | No | No | No |
| Apple | FaceTime (E2EE by default) | No | No | No |
| Meta | Messenger | Yes (when not E2EE) | Yes (when not E2EE) | Yes (when not E2EE) |
| Meta | WhatsApp (E2EE by default) | Yes (on profile & group photos, user reports) | Yes (on user reports) | Yes (on profile & group photos, user reports) |
| Microsoft | OneDrive | Yes (when material is shared) | No | No |
| Microsoft | Skype/Teams | Yes (when not E2EE) | Yes (When not E2EE) | No |
| Google | Drive | Yes | Yes | Yes |
| Google | Messages | No | No | No |
| Google | Meet | No | No | No |
| Google | Chat | Yes | No | No |
| Google | Gmail | Yes | No | No |
| Google | Google Photos | Yes | Yes | Yes |

While the results from the transparency notices offer insights and evidence of mitigations in place by 'large-scale' international RES providers (>50,000 employees), they do not provide sufficient data to establish the level of existing mitigations for small, medium, or large Australian based RES or DIS, or an estimate of business-as-usual costs.

### ii. The size, complexity and variability of the services covered.

A proportion of RES and DIS providers in scope of the standards are international companies who provide multiple services (messaging, social media etc), and have large operating costs and high revenue. These international services are in some

cases likely to have implemented, or be in the process of implementing systems, processes, and technologies in response to online safety regulations in other jurisdictions as well as their own voluntary commitments. While these are not costed in regulatory burden estimates, it is important to highlight how these vary from Australian RES and DIS providers. High-risk Australian RES and DIS, particularly small to medium sized businesses with a lower revenue and resources, are less likely to have existing mitigations and may be disproportionately impacted by regulatory costs, in particular the more onerous obligations such as deploying technologies to detect known child sexual abuse or pro-terror material. Other apps or websites with different business models including most Australian RES and DIS providers are also in scope (due to the provision of a website or app), however many are unlikely to have any compliance obligations and therefore limited, to no regulatory costs.

#### iii. The risk classification of services, and the implementation requirements, vary – impacting obligations and regulatory burden costs.

Services that are deemed to have a higher risk for access, production, and distribution of class 1A and 1B material have more obligations under the Option 3 (standards) and higher regulatory costs. For example, while there are meaningful obligations on high-impact websites and apps DIS, most DIS (for example general purpose news, educational, health or retail websites) will not have any meaningful obligations under the DIS standard and therefore no, or limited, regulatory costs[30]. Option 3 (standards) also provide that in some instances services are not required to implement systems and technologies if they can demonstrate that it is not technically feasible or reasonably practicable to do so, or where it would result in a systemic weakness or vulnerability into the service, or in the case of an end-to-end encrypted service would result in a new decryption capability or render methods of encryption used in the service less effective.

These elements of the Option 3 (standards) provide flexibility for providers and consider what is reasonably practicable, which may include considerations such as cost. Where it is not technically feasible or reasonably practicable to

[30] DIS as defined in the Act includes a wide variety of unique services and will include most apps and websites that can be accessed by end-users in Australia. This includes for example grocery and retail websites, websites containing contact and service information for small businesses such as cafes, hairdressers and plumbers, apps offered by medical providers to allow patients to access x-ray imagery, information apps such as train or bus timetable apps, newspaper websites, as well as websites aimed at providing educational, information and entertainment content to Australian end-users. Most these services will have no obligations given they present low risks.

implement a system or a technology service providers must undertake appropriate alternative action.

#### iv. Costs are highly variable and depend on the obligations.

The key obligations in the Option 3 (standards) and Option 2 (drafted codes) will have different costs associated with them. Certain obligations, specifically those relating to the detection and CSAM and pro-terror material removal (involving the deployment of systems, processes, and technology to proactively detect) are likely to incur the greatest amount of cost to high-risk service providers. The estimates for content moderation in the regulatory burden estimates below (section 4.2) are illustrative only of the costs likely to be incurred in deploying the requirements. They are also considered to significantly overestimate the costs to Australian RES and DIS providers in scope of Option 3 (standards). This is because the UK OSA applies to a much broader scope of material than class 1A and class 1B (the subject of the policy options (please refer to Annexure B Table 12 and 13 for a breakdown of compliance costs).

#### v. Technology to proactively detect known material is available at no cost.

Several "hash matching" tools are available free which can be deployed to assist service providers meet the relevant requirements in the Option 3 (standards) to detect and remove certain known material. These tools create a unique digital signature (known as a 'hash') of an image which is then compared against signatures (hashes) of other images to find copies of the same image. The following hashing tools are currently freely available:

- Microsoft and Dartmouth College's PhotoDNA (eSafety Commissioner, 2022)
- Facebook's open-source photo and video matching technology (Davis & Rosen, 2019)
- Google's hashing tools for videos, and tools for detection of new images (Google, 2024)

Although freely available technology means there is no build cost, there are still implementation, support, and maintenance costs to be considered in the adoption of this technology. Companies may also choose to deploy trust and safety personnel within a service or engage external content moderation services. These costs vary based on the solution, the volume of content being scanned and the complexity and size and the service the tools are being built for. These costs may disproportionately impact small to medium Australian RES and DIS providers that are assessed to be

Tier 1 or in the pre-assessed risk categories (please refer to Annexure D – Risk categories for RES and DIS providers).

## 4.2. Regulatory burden estimates

Using the methodology described in section 4.1.2, the likely net benefits of the policy options is estimated below. The assumptions underpinning the following regulatory burden estimates and a breakdown of individual compliance obligations and costs are provided in Annexure B.

### 4.2.1. Option 1 (maintain status quo)

Option 1 (maintain the status quo) would require no change by RES and DIS providers to their approach to management of the risks associated with class 1A and class 1B material on their services. There is no regulatory burden for community organisations or individuals. Option 1 therefore has a zero estimated regulatory burden cost, as it represents the business-as-usual case[31] and does not have any additional administrative or substantive compliance or delay costs. It does not introduce any new regulatory costs to businesses, communities, or individuals.

Option 1 would provide no community safeguards that would curb the production, distribution, and consumption of class 1A and 1B material online. There would continue to be significant economic, health and social impact through harms to individuals and community due to the higher risk of class 1A and class 1B on these services.

**Table 2: Total Regulatory burden estimate table – Option 1 (maintain status quo)**

**Total regulatory costs at end of 10-year appraisal period (from business as usual)**

| Change in costs ($ million) | Business | Community Organisations | Individuals | Total change in costs |
|---|---|---|---|---|
| Total, by sector | $0 | $0 | $0 | $0 |

[31] Business as usual costs being excluded from the Government Regulatory Burden Measurement framework, which is designed to measure regulatory burden over and above what a normally efficient business (defined as an entity that handles its regulatory tasks no better or worse than another) would pay in the absence of the regulation.

### 4.2.2. Option 2 (industry co-regulation)

Option 2 (industry co-regulation) would require the development of RES and DIS industry codes for class 1A and 1B online material which are able to be registered by the eSafety Commissioner. While draft industry codes have been developed, these were not registered as they did not provide appropriate community safeguards. There is no meaningful regulatory cost burden for community organisations or individuals for Option 2. This is because community organisations or individuals are unlikely to operate a RES or a DIS that incurs obligations under the DIS standard (i.e. make available high-impact content).

The regulatory burden costs for Option 2 below represent a proportion of the UK OSA estimates. The variation in compliance obligations was determined via a qualitative assessment of the drafted industry codes (based on substantive requirements themselves in addition to enforceability and scope) and the UK OSA comparative obligations. A proportion was assigned to each compliance obligation for Option 2 and costs were then adjusted based on this estimate.

It is estimated that at the end of the 10-year appraisal period the total regulatory burden cost to businesses in scope of Option 2 (including compound growth on businesses in scope) will be $135 million. It is estimated the average total regulatory cost burden per business at the end of the 10-year period will be $70,000. This is the average cost and does not differentiate the costs based on risk/obligations of the service, its size (turnover/employees), capital or labour costs. **A breakdown of individual costs is provided in Annexure B – Table 12.**

**Table 3: Total Regulatory burden estimate table – Option 2 (industry co-regulation)**

**Total regulatory costs at end of 10-year appraisal period (from business as usual)**

| Change in costs ($ million) (rounded) | Business | Community Organisations | Individuals | Total change in costs |
|---|---|---|---|---|
| Total, by sector | $135 | $0 | $0 | $135 |

It is estimated that at the end of a 10-year appraisal period the total **annual** regulatory burden cost to Australian businesses with obligations under Option 2 (including compound growth) will be $14 million.

**Table 4: Annual Regulatory burden estimate table – Option 2 (industry co-regulation)**

Total annual regulatory costs (from business as usual)

| Change in costs ($ million) (rounded) | Business | Community Organisations | Individuals | Total change in costs |
|---|---|---|---|---|
| Total, by sector | $14 | $0 | $0 | $14 |

There are several obligations under Option 2 that are not costed in the regulatory burden above, due to the absence of available data to obtain these estimates (i.e., these were not obligations under the UK OSA). Some of these provisions in the drafted codes include requirements for safety features and settings, trust, and safety function (adequate personnel/resources) and ensuring that eSafety information is available to end-users.

### 4.2.3. Option 3 (direct regulation)

Option 3 (direct regulation) in the form of industry standards will result in new regulatory costs on businesses that have obligations under the standards. There is no regulatory burden for community organisations or individuals. As above, there is no meaningful regulatory cost burden for community organisations or individuals for Option 2. This is because community organisations or individuals are unlikely to operate a RES or a DIS that incurs obligations under the DIS standard (i.e., make available high-impact content).

Regulatory burden costs associated with direct regulation are difficult to quantify with any precision. Compliance costs can be expected such as the costs of putting in place new technologies, systems, and processes to meet regulatory requirements, possible human content moderation and evolving system requirements, as well as administrative compliance costs such as the cost of reporting on compliance, conducting risk assessments and keeping records. Costs may also be incurred by providers in providing mechanisms for users to report complaints or breaches and updating, enforcing, and making available terms of service.

The RES standards place obligations on the providers of email, private messaging, chat services and other communication services. While all websites and apps not falling within other industry sections subject to either a code or standard is a DIS,

most DIS will not be subject to specific obligations under the DIS standard. However, high-impact websites (such as pornography or 'gore' sites), file and photo storage services, certain online services with generative AI capability, and platforms which distribute open-source machine learning models will have obligations. These online service providers will likely need to deploy technology and/or allocate more personnel, services, or time to comply with the standards. However, some providers of RES and DIS with effective online safety measures may already be compliant with key obligations.

The distribution of costs is also difficult to determine based on the size of the RES and DIS (according to either their turnover or available resources). Based on the baseline data for services in scope, it is assessed that most of the Australian RES and DIS will be micro (0-4 employees) to small businesses (5-19 employees). While the distribution of Australian RES and DIS within the different risk categories is unclear (see Annexure D for risk categories), it is expected that most of the micro - small service providers will also have limited to no, obligations under the Standards. However, if these services are risk classified as Tier 1 or are a pre-assessed RES or DIS, they will have more significant obligations and therefore likely to have a higher regulatory burden. Key obligations on these RES and DIS are however still subject to limitations such that they do not apply where the requirement would not be technically feasible or reasonably practicable, or where it would introduce a systemic weakness or vulnerability. In addition, many obligations include 'appropriate' in them which enables compliance to consider proportionality and the reach of a service.

The regulatory burden costs for Option 3 below represent a proportion of the UK OSA estimates. The variation in compliance obligations was determined via a qualitative assessment of the drafted industry codes (based on substantive requirements themselves in addition to enforceability and scope) and the UK OSA comparative obligations. A proportion was assigned to each compliance obligation for Option 3 and costs were then adjusted based on this estimate.[32]

It is estimated that at the end of 10-year appraisal period the total costs to RES and DIS providers for Option 3 is $212 million (including compound growth on businesses in scope). It is estimated the average regulatory cost burden per business at the end of the 10-year period will be $100,000. This is the average cost

[32] For example: for 'undertaking content moderation' is estimated to only cover 20 % of the obligations in Option 3 – Standards. The compliance costs for Option 3 were then adjusted for this obligation by 0.20 % of the total costs.

and does not differentiate the costs based on risk/obligations of the service, its size (turnover or capitalisation or number of employees. It is expected that the cost burden will be mostly incurred in the first year for high-risk services who have obligations requiring the implementation of systems and/or technologies and who have no existing mitigations. **A breakdown of individual costs is provided in Annexure B – Table 13.**

**Table 5: Total Regulatory burden estimate table – Option 3 (direct regulation)**

**Total regulatory costs at end of 10-year appraisal period (from business as usual)**

| Change in costs ($ million) (rounded) | Business | Community Organisations | Individuals | Total change in costs |
|---|---|---|---|---|
| Total, by sector | $212 | $0 | $0 | $212 |

The annual estimated regulatory burden costs to businesses in scope of option 3 (including compound growth) is $21 million at the end of 10-year appraisal period.

**Table 6: Annual Regulatory burden estimate table – Option 3 (direct regulation)**

Total annual regulatory costs (from business as usual)

| Change in costs ($ million) (rounded) | Business | Community Organisations | Individuals | Total change in costs |
|---|---|---|---|---|
| Total, by sector | $21 | $0 | $0 | $21 |

There are several obligations under Option 3 that are not costed in the regulatory burden above, due to the absence of available data to obtain these estimates (i.e., these were not obligations under the UK OSA). Some of these provisions in the standards include requirements for safety features and settings, resourcing trust and safety, development, and investment program[33] and ensuring that eSafety information is available to end-users. Further detail of these obligations is provided in Table 7.

Comparative to Option 2, which did not provide adequate safeguards, Option 3 will create a safer online environment for individuals and the community, and further

[33] Only applies to some high risk RES and DIS with monthly active end users over 1,000,000 in the previous calendar year.

protection from harms stemming from access, exposure to Class 1A and 1B material. It will also strengthen transparency of, and accountability for this type of material by RES and DIS providers. The following table provides an overview of the compliance measures under Option 3 and how they are expected to reduce harms/achieve positive outcomes.

**Table 7 - Option 3 – Examples of Compliance Obligations and expected harm reduction outcomes.**

| Obligation | Action Required | How the measure will result in harm reduction/outcomes |
|---|---|---|
| **Providing a Mechanism for reports and complaints on material and breaches of terms of use**<br>**Responding to breaches and terms of use** | Provide a tool, to enable end users to make reports and complaints.<br>Take appropriate action to prevent further access to material and minimise further breaches.<br>Remove material as soon as practicable and take appropriate action – where not technically feasible or reasonably practicable.<br>*Applicable to certain categories of RES and DIS* | BOSE Transparency reports indicate that user reporting features are commonly implemented by services but that these vary in their accessibility for users, and in services' responses. (eSafety Commissioner, 2022)<br>The obligations will ensure that reporting mechanisms are in place that ensure end users can make a complaint or report and that material is removed. This is expected to lead to a reduction in the circulation of harmful material. |

| Obligation | Action Required | How the measure will result in harm reduction/outcomes |
|---|---|---|
| **Detecting and removing known CSAM and PTM**[34] | Must implement (*where technically feasible and reasonably practicable*) appropriate systems, processes, and technologies to detect and remove known CSAM and PTM on their service – high risk services only. | In 2022, NCMEC's CyberTipline received more than 32 million reports of suspected child sexual exploitation. Reports of CSAM discovered online was 90% higher in 2020 than 2019. (Fitzsimmons, 2021)<br>Research on social media[35] has shown that content moderation can curb online harm and that if platforms that do not moderate harmful content can generate more material that can lead to exponential growth. (Rizoiu & Schneider, 2023). Detection of known CSAM and pro-terror content is part of content moderation.<br>Proactive detection and removal of CSAM and PTM is expected to lead to reduction in harms relating to the access, production, and distribution of this type of material. It will also assist in increasing detections of hashed material that has been distributed in other jurisdictions (i.e., through NCMEC) and curb growth of this material online. |
| **Disrupting and Deterring CSAM and PTM** | Must implement systems and processes, and if it is appropriate technology to disrupt and deter CSAM and PTM on their service-high risk services only. | |
| **Safety Features and Settings (including Resourcing)** | Assess safety features before making a material change to service, obtain user registration details and provide info on safety tools and settings. | Providing information on safety tools and settings to users that are accessible and easy to use will afford greater protections to end-users, particularly children. This also includes enabling users to block their status, ensure privacy by default settings for under sixteen, and prevent adults from contacting children without parental/carer consent. This ensures Safety by Design Principles are considered across platforms when there is a material change. |

[34] Pro terror material.

[35] Research was undertaken on social media which is not a service in scope of the standards, however it does reflect the reduction in harms which are applicable for all content moderation, including RES and DIS providers.

| Obligation | Action Required | How the measure will result in harm reduction/outcomes |
|---|---|---|
| Development Program | Must establish and implement a program of investment and development activities- (for RES with 1 million MAU in past year and some high risk DIS). | Increased investment in trust and safety systems, processes and technologies would see a reduction in online harm. Better information and intelligence sharing relationships between service providers, government and non-government organisations will also reduce harms, through proactive identification of new risks, emerging technologies/harms and solutions. |
| eSafety Information available to end-users | Dedicated location for information available to end-users. | According to ACCCE research - 51% of participants did not know what they could do to keep children safe from online child sexual exploitation and only 52% of participants talk to their children about online safety. (The Australian Centre to Counter Child Exploitation, 2020) Information provided to Australian end users about the risks and prevalence of online harms on platforms and e-safety initiatives, will mitigate some of the online harms through increased education and prevention. An obligation to put this information in a dedicated location will ensure that end users have ready access to information that keeps them informed on eSafety information to enhance online safety. |
| Risk Assessments | Require in-scope services and platforms to undertake risk assessments where there has been a material change to their service that increases the risk of class 1A or 1B material on their services. | Many platforms already conduct risk assessments; however, there will be some that do not, and these assessments could result in more or better targeted safety measures such as content moderation leading to greater harm mitigation. |

| Obligation | Action Required | How the measure will result in harm reduction/outcomes |
|---|---|---|
| Reporting to eSafety | Notify eSafety of new features, technical feasibility, and outcomes of development programs. Compliance reports may be required (on request of eSafety Commissioner) | Ensures Safety by Design Principles are considered with the implementation of any new features and assessment of any increase in risk for Class 1A and 1B material and adjustment of compliance obligations. Ensures industry accountability with investment in development programs and technical feasibility reports. Enables eSafety to work with industry to minimise emerging risks and reduce online harms to end-users. |

## 4.3. Estimating quantifiable harms

It is noted that any attempt to estimate the monetary costs of abuse may seem reductive to victim-survivors, their families, and others in the community. This analysis is not intended to diminish the terrible impacts experienced by victim-survivors and their families in any way – any financial quantification of harms can never represent the considerable and unmeasurable human costs of abuse.

The technical and research resources required to conduct a full cost benefit analysis and the timeframe required for such were prohibitive and could not be achieved within the scope of the requirements for the introduction of the standards. Estimation of the overall benefits of the options is therefore difficult to determine as it is not possible to develop a precise valuation of the reduction in harm comparative between each option.

There is no available research or data quantifying directly the harms from class 1A and class 1B material on RES and DIS. The International Centre for Missing and Exploited Children (ICMEC) has recently opened applications for Australian academics to submit their interest in conducting new research into the *economic* consequences and impacts of child sexual exploitation, particularly facilitated online. There are no current Australian studies that have quantified the specific economic costs resulting from exposure to CSAM online, or other harmful material.

The quantified harms in this section therefore derive predominately from the analysis of international studies where the costs of harm from child sexual abuse and online child sexual abuse were estimated. These are used to indirectly provide

an estimate and basis of the likely costs of online child sexual abuse and child sexual abuse which could reasonably be expected in Australia. Costs stipulated in these studies (due to historical nature) have been adjusted for inflation and converted to Australian dollars.

### 4.3.1. Online Child Sexual Abuse

The 2019 impact assessment undertaken for the UK OSA estimated the proportion of contact child sexual abuse[36] with an 'online element' to be 20.1% of all child sexual abuse offending in the UK. It is estimated that child sexual abuse with an online element costs **A$2.1 billion**[37] per year in the UK in 2023.

**Table 8 - Estimated annual cost of online contact child sexual abuse (in AUD and adjusted for inflation)[38] – UK OSA**

| Harm | Estimated UK annual cost | Proportion online (UK OSA estimate) | Annual AUD cost with online elements |
|---|---|---|---|
| Contact child sexual abuse | A$10.7 billion | 20.1 % | A$2.1 billion |

This figure does not provide an estimate of the cost in Australia and reflects the UK findings only.

Two further studies have estimated the costs of 'child sexual abuse' more broadly (not online specific) from the United States (2018) [39] and United Kingdom (2014)[40]. These studies estimate the annual cost of child sexual abuse in these jurisdictions to be between A$8.2 and A$18.4 billion. While encompassing a much broader array of offending and variability in their scope, definitions, population, methodology, sample size, and timeframes – these studies highlight the immense costs associated from this type of offending.

If the online component of child sexual abuse is estimated to be 20 per cent of all child sexual abuse offending (Table 9), is applied to these broader studies the

[36] Child sexual abuse can comprise of contact activities /physical contact (e.g., rape, unwanted touching) and non-contact - without physical contact (e.g. exhibitionism, exposure to pornography, verbal sexual harassment, distribution of intimate pictures against one's will).
[37] Source figures have been adjusted for Inflation in country of origin (2023) and currency conversion to AUD.
[38] Source figures have been adjusted for Inflation in country of origin (2023) and currency conversion to AUD.
[39] Letourneau, E.J., Brown, D.S., Fang, X., Hassan, A., & Mercy, J.A. (2018). The economic burden of child sexual abuse in the United States. Child Abuse & Neglect
[40] Saied-Tessier, A. (2014). Estimating the costs of child sexual abuse in the UK. NSPCC

costs are broadly comparable to the annual estimates of child sexual abuse with an online element in the UK OSA.

**Table 9 - Annual cost of child sexual abuse in the United Kingdom and United States**

| Study | Estimated annual cost of 'child sexual abuse' (adjusted inflation/AUD) |
|---|---|
| United Kingdom (2014) | A$8.2 billion |
| United States (2018) | $A18.4 billion |

These studies cannot be directly applied to the Australian environment without adjustment for differences in health care, welfare, job markets, offence reporting, criminal justice, and education systems. However, based on prevalence rates of child sexual abuse in Australia and emerging evidence on the prevalence of CSA facilitated at least in part online, should economic analysis of the impact of online child sexual abuse be undertaken in Australia, it is likely to reveal costs of a similar and significant magnitude (but potentially adjusted to population size). While these costs are significant, it is reiterated that the burden of 'online' child sexual abuse is unlikely however to all be linked just to RES and DIS and the exact proportion that could be attributed to RES and DIS cannot be estimated. Refer to Appendix B Table 14 for further breakdown of these studies.

Further, given that a substantial proportion of child sexual abuse is not reported,[41] including that which occurs online, it is highly likely that these figures understate the economic costs to government, community, and individuals within these jurisdictions. This also does not capture the costs that would reasonably be incurred on individuals, community and government resulting from other harmful material such as pro-terror and extreme violence being accessed, produced, and distributed on RES and DIS.

[41] Both studies state that their estimates are likely conservative – for example, the United States (2018) study is based on data from child protection agencies and notes that not all cases of child abuse are reported to authorities.

## 4.4. Summary of costs and benefits

In summary[42]:

- Option 1 (maintain the status quo) has no regulatory cost burden to businesses, individuals, and community organisations and would provide no community safeguards that would curb the production, distribution, and consumption of class 1A and 1B material online. There would continue to be significant economic, health and social cost through harms to individuals and community due to the higher risk of class 1A and class 1B on these services.

- Option 2 (industry co-regulation) has some regulatory burden costs to businesses in scope, although not as significant as Option 3 - due to the draft codes having less obligations than the standards and reduced enforceability of key obligations. There would be some additional safeguards that would curb the production, distribution, and consumption of class 1A and 1B material online, but there would continue to be significant economic, health and social costs through harms to individuals and community due to the higher risk of class 1A and class 1B on these services; and

- Option 3 (direct regulation) has the most significant regulatory burden costs for businesses in scope. Option 3 provides the highest net benefit in harm reduction through the provision of safeguards to curb the production, distribution, and consumption of class 1A and 1B material online, and is expected to have a greater impact on reducing the economic, health and social impact to individuals and community by reducing the risk of class 1A and class 1B on those services covered by the standards.

**Option 3 (direct regulation) is estimated to have the greatest annual net benefit** while a benefit-cost ratio cannot be quantified (due to the absence of data on the harm/cost mitigations for each policy option) it is assessed that the implementation of the Standards will highly likely lead to a reduction in the risk and growth of class 1A and class 1B on RES and DIS services, which will have a

[42] Noting that as previously outlined, regulatory cost estimates for Options 2 and 3 are almost certainly overestimated, particularly costs for obligations that involve content moderation activity (detect and remove/disrupt and deter provisions). These specific provisions are also subject to technical feasibility exemptions, with the level of obligations also proportionate and appropriate to the level of risk of class 1A material being on a service. As already highlighted, though not provided, costs will be borne differentially by different providers depending on their size, risk tier, and existing mitigations.

direct reduction in harms. Option 3 will provide a cost benefit to individuals, community, and government through a reduction in harms and associated economic, health and social impacts. Mitigation of these harms and associated costs (both tangible and intangible) is why Option 3 is considered to provide the greatest annual net benefit of the policy options.

# 5. Who did you consult and how did you incorporate their feedback?

This chapter outlines the consultation undertaken to develop the standards, the principal views of stakeholders (including areas of agreement and disagreement), and how the preferred option has been modified to take account of stakeholder views.

## 5.1. Details of consultation

In November 2023 the eSafety Commissioner invited submissions[43] from the online industry, advocacy groups, other stakeholders, and the public on the two draft industry standards for RES and DIS under the Online Safety Act 2021. This engagement followed industry associations' 12 months plus engagement with these stakeholders in the development of the draft RES and DIS codes.

eSafety's consultation was an important part of the process to better understand the impact of proposed obligations on industry as well as the concerns of advocacy groups. Given the large scope of providers who could be categorised as RES or DIS it was important to obtain feedback from providers of different size, service offerings, and risk profiles to understand the impact of the standards across a broad range of providers.

Additionally, it was valuable to receive submissions from stakeholders across different civil society groups such as child rights and privacy groups. Consultation invited concerns to be raised, with feedback being considered and addressed in the final standards. The transparency and public scrutiny of the draft standards contributed to final standards that are measured and balanced.

[43] Industry standards – public consultation | eSafety Commissioner

To assist stakeholders and interested parties to comment during the consultation, a discussion paper[44] and fact sheets[45] were released alongside exposure drafts of the Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024[46] and the Online Safety (Designated Internet Services — Class 1A and Class 1B Material) Industry Standard 2024.[47]

The consultation was publicised via media release and social media, and emails were sent to 200 Australian and international stakeholders including civil society human and children's rights groups, generative AI experts, relevant government bodies and key industry associations advising them of the consultation. The consultation period was formally open for 31 days; however, it was made clear that extensions were available to account for the limitation that the consultation period was not if some stakeholders would have liked. eSafety granted around 20 extensions to ensure submitters had adequate time to provide a considered and meaningful submission. All parties that requested an extension were granted one.

The discussion paper set out the legislative framework for the standards; outlined eSafety's overarching approach to the standards; and included questions on key elements of the standards. The complexity and breadth of the draft standards could have been a potential barrier for industry, stakeholders, and the public to provide suitable feedback on the standards. Accordingly, the questions in the discussion paper were designed to assist and guide direct feedback on critical issues - but it was made clear the questions were not intended to limit the scope of submissions. The discussion paper also specifically requested views on the estimated costs for RES and DIS providers of compliance with the relevant standard, and the impact of compliance costs on potential new entrants to these sections of the online industry. However, as set out above, this information was not forthcoming.

---

[44] Discussion Paper: Draft Online Safety (Relevant Electronic Services - Class 1A and 1B Materia) Industry Standard 2024 and Draft Online Safety (Designated Internet Services - Class 1A and 1B Material) Industry Standard 2024

[45] Fact sheet: Draft Online Safety (Designated Internet Services – Class 1A and Class 1B Material) Industry Standard 2024 and Fact sheet: Draft Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024

[46] Draft Online Safety (Relevant Electronic Services - Class 1A and Class 1B Material) Industry Standard 2024.pdf

[47] Draft Online Safety (Designated Internet Services-Class 1A and Class 1B Material) Industry Standard 2024.pdf

In addition, to provide stakeholders with further opportunities to provide feedback on the draft standards, eSafety held two round-table discussions with key stakeholder groups in December 2023.

The first roundtable included representatives from industry associations and service providers from the two relevant industry sections. The second roundtable involved stakeholders from different civil society organisations including children's rights and digital rights groups, and academics. The roundtables were an important contribution to informing the development of the draft standards and an efficient way to obtain direct comments from key industry representatives and civil society groups on the draft standards and for eSafety to clarify certain points.

eSafety also met with industry, organisations, and government agencies before, during and after the consultation period to discuss the draft standards. This included working closely with the Department of Industry, Science and Resources to avoid an inconsistent approach across government on AI-related regulation and focusing the DIS standard on targeted obligations for high-risk consumer facing services.

In the lead up to the formal consultation period for the standards, eSafety engaged in significant consultation on generative AI online safety issues. In August 2023, eSafety published a position statement on generative AI as part of our Tech Trends and challenges workstream. The statement was informed by extensive consultation with a range of domestic and international AI experts, representatives of the eSafety Youth Council and Trusted eSafety Providers (TEPs) program, as well as feedback from inter-departmental colleagues (including the OAIC). eSafety then engaged in targeted consultation with generative AI online safety experts prior to release the draft Standards.

Topics covered in the consultation on the RES standard included:

- The role of risk assessments to reduce the risks of class 1A and class 1B material being generated, posted, stored, or distributed. The draft standards propose that providers of certain services self-assess their risk to identify their risk tier and consequent legal obligations.
- The appropriateness and effectiveness of the technical feasibility exception to the obligation to detect and remove known child sexual abuse material and pro-terror material.

- Whether there are any limitations which would prevent certain service providers from deploying systems, processes, and technologies to disrupt and deter child sexual abuse material and pro-terror material on RES and if so, how they might be overcome.
- Whether stakeholders agreed with the 'monthly active user threshold' for the investment obligation, or whether there are other appropriate thresholds that should be considered to ensure the obligation is proportionate to the size and reach of RES.
- Whether end-user reporting requirements are workable for RES providers, or if there are practical barriers to implementation.
- Whether the requirement on certain RES to respond to reports of class 1A and class 1B material on their service should be limited to a requirement to take 'appropriate action'.

Topics covered in the consultation on the DIS standard included:

- Whether the risk categories are sufficiently clear for DIS providers to identify which category they fall within and therefore what obligations apply, as well as the benefits and/or challenges of the categories proposed.
- Whether the provisions regarding generative AI are appropriate, meaningful, and targeted effectively to achieve the desired result, and whether there are specific challenges to deploying measures in a generative AI context.
- In relation to model distribution platform, whether the proposed obligations provide appropriate safeguards, and any specific challenges to deploying these measures.
- In relation to relevant enterprise providers, whether proposed obligations provide appropriate safeguards, and any specific challenges to deploying these measures.
- Whether the technical feasibility exception to the obligation to detect and remove known child sexual abuse material and pro-terror material is appropriate and whether the exception impacts the effectiveness of the obligation.
- Whether the monthly active user threshold for investment requirements is appropriate.

## 5.2. Principal views of the stakeholders

The written submissions received by eSafety on the draft standards were published on the eSafety website in February 2024[48]. These were redacted to remove personal or sensitive information (such as physical addresses, telephone numbers and email addresses) and information identified as confidential.

It is important to note that not all submissions commented on every element of the standards and many focused on the standard that would apply to them.

The major themes identified in the submissions included:

- Definitional issues
- Detection and removal of pro-terror material
- The application of the technical feasibility exception
- Impact on end-to-end encrypted services
- child protection
- generative AI service categories
- risk assessments

An outline of the principal views of stakeholders is discussed below.

### 5.2.1. Areas of agreement and difference

eSafety closely considered the submissions received and what amendments should be made, including amendments to provide greater certainty to both industry participants and end-users. Opinions on the draft standards varied, in part due to the wide scope of the RES and DIS standards themselves, but also because of wide variance in the interests and positions of stakeholders impacted by the standards with different views expressed by each of digital rights advocates; privacy advocates; child protection groups; and industry associations/service providers, with each advocating in line with their primary interest.

[48] Industry standards – public consultation | eSafety Commissioner

In summary, the key issues raised which were often the most common areas of agreement and differences included:

#### 5.2.1.1. Feedback from civil society groups

Child protection groups and digital rights groups often had a strong divergence of views on the same compliance measure.

Child/human rights groups were supportive of direct regulation, in addition to providing feedback on the following.

- The draft RES standard defined 'young Australian child' and 'Australian child'. Child rights groups were concerned that this appears to create a ceiling age of 16 for certain protections in the draft standards and does not align with international laws definition of a child. Digital rights groups did not raise this as an issue, however, some individuals/academics did.
- The draft standards had obligations on certain providers to implement development programs. Child/human rights advocates recommended strengthening this measure by amending the provision to ensure they are 'genuine' development programs by making service providers commit to this obligation in good faith to mitigate any tokenistic measures.
- The technical feasibility provision in the draft standards specifies certain matters to be considered when assessing what is technically feasible or reasonably practicable including the expected financial cost to the provider of taking the action, and whether that is reasonable for the provider to incur having regard to the extent of the risk. Child rights/human rights groups recommended stronger wording as the exception may leave platforms with limited responsibility to prioritise child safety.
- That the generative AI categories capture the right platforms and services to address the risks of synthetic child sexual exploitation material.

Privacy/digital rights groups were generally supportive of direct regulation to prevent child sexual abuse and other illegal material; however, they expressed strong concerns re the potential erosion of privacy and the impact on end-to-end encrypted services.

- The key concern of this group was the absence of an explicit carve out for end-to-end encrypted services from the requirements to implement a system, processes, and technologies to detect and remove certain known material. Large service providers and their industry associations as well as

individuals submitted similar concerns that privacy and security was not referenced in the technical feasibility exception.

- An additional concern was that the technical feasibility exception was only applicable to the 'detect and remove' child sexual abuse material and pro-terror obligation and did not extend to the 'disrupt and deter' measure that is required if a provider is unable to meet the detect and remove obligation. Several service providers and their industry associations also shared this concern.

#### 5.2.1.2. Feedback from industry

RES and DIS providers expressed concerns regarding methodology, wording and definitional concerns, technical feasibility issues, risk assessments, and end-to end-encryption. Providers were asked about the estimated cost of adoption of the draft standards however this information was not provided.

There was feedback from some RES and DIS providers on the categorisation of generative AI services. Feedback from these providers varied with some proposing removing the generative AI service categories entirely from the DIS standard and waiting for broader government reforms on AI with others proposing refining the categories through amendments and other supporting the supply chain categorisations. Feedback also included requests for greater clarity on the services intended to be caught by generative AI and narrowing or expanding definitions. Industry associations had a similar sentiment and provided like feedback.

Some service providers also felt strongly about the compliance measures required for certain generative AI services. Several large providers and industry associations submitted that the obligations on model distribution platforms and generative AI model developers were disproportionate and not feasible.

RES and DIS industry associations and some providers expressed concerns about the 'predominant functionality' test to determine whether a service is covered by one of the standards and requested alignment with the predominant purpose test in the Head Terms of the registered codes.

#### 5.2.1.3. Where feedback aligned across interest groups

Digital privacy rights organisations, some service providers and industry associations expressed concerns that the requirement to proactively detect pro-terror materials should be amended to clarify that the need to comply with this

obligation is only to the extent that material has been sent or shared with another person and not material stored in an 'inert' state.

An industry association requested clarification of the risk assessment requirements to include further matters to be considered when determining a risk profile such as existing mitigations. A civil society organisation and a large service provider also made a similar suggestion.

## 5.3. Revision of the standards to take into account the feedback received

The feedback received from stakeholders helped shaped the development of our most viable option – Option 3. The draft standards were amended and finalised after considering the feedback from industry participants, industry associations, government agencies, civil society organisations and the public. eSafety closely considered the submissions received and what amendments should be made, including amendments to provide greater certainty to both industry participants and end-users. Where feedback received during consultation period was not incorporated into the final RES and/or DIS standard, this was based on consideration of the policy objective, eSafety's powers under the Act, the scope of the standards, and evidence provided in the submissions - including the effectiveness and workability of drafting and the likely beneficial contribution of the amendment to the objectives of the relevant standard and provision.

Approximately 200 separate issues were identified and considered by eSafety from the feedback received, and the draft standards were amended significantly. Some of the key changes to the standards included:

- Amending the test in the DIS Standard to determine which code or standard a service must comply with from a 'predominant functionality' test to 'predominant purpose' test, and changing purpose to functionality in some DIS category definitions.
- Specifying that there is no requirement to build a systemic weakness or vulnerability into end-to-end encryption; or build a new decryption capability in relation to an encrypted service; or render methods of encryption less effective.

- Limiting detection and removal requirements in relation to pro-terror material 'at rest' (i.e., in inert spaces such as file/photo storage or emails in draft form).
- Clarifying how 'appropriate' is to be interpreted to ensure that matters like proportionality and potential harms are considered in how a provider complies with obligations.
- Removing the open and closed RES categories and creating a new general definition of 'communication RES' to cover both closed and open communication RES.
- Removing dating services from the obligation to detect and remove pro-terror material in the RES standard. This provision now applies to Tier 1 RES, communications RES, and gaming services with communications functionality.
- Adding a requirement that users be allowed to request review of the outcomes of their complaints regarding material has been added to report handling requirements in the RES standard.
- Clarifying the scope of categories of generative AI services to address uncertainty.
- Limiting, at this stage, the obligations to be placed on upstream model developers while the eco-system for generative AI services develops and broader regulation is considered. High-risk consumer facing generative AI services and model distribution platforms continue to be covered, consistent with feedback from AI child safety experts.
- Clarifying that a high impact generate AI DIS does not include a service which has guardrails and controls in place such that there is an immaterial risk that end-users can generate synthetic high impact material.
- Removing some obligations applying to model distribution platforms and clarifying how obligations may apply. The category name was also changed from 'machine learning model platform service' to 'model distribution platform'.
- Deeming Enterprise DIS providers to be Tier 3 (low risk), and so removing requirements specific to enterprise DIS throughout the DIS Standard.

# 6. What is the best option from those you have considered and how will it be implemented?

**In this chapter, the recommended option and how it was identified is discussed, along with the approach to implementation; the implementation challenges, implementation risks and their management; and the anticipated implementation timeline and transitional arrangements.**

## 6.1. How we identified the recommended option

Building on more than two years of consultation by industry associations on development of the draft industry codes and the feedback received via eSafety's consultation on the draft standards, eSafety has identified Option 3 (amended in response to feedback) as the best option to provide an appropriate protection in respect of class 1A and class 1B material on RES and DIS.

Consistent with the Australian Government Guide to Policy Impact Analysis, eSafety considered policy options against:

- the quantitative cost-benefits
- qualitative benefits; and
- the feedback from consultation

to establish the most effective, appropriate, and efficient option which had the greatest net benefit for Australia.

Using the guiding OIA principle that the best option is that with the highest net benefit and is the most effective, appropriate, and efficient option, we determined that Option 3 – registration by the eSafety Commissioner of the final RES and DIS standard has the highest net benefit for Australia and is our recommendation.

As highlighted in chapter 4 and in Annexure B several assumptions were made in determining the likely net benefit of each option.

The proposals and evidence provided throughout this document have given some weight to the Government's view that industry providers need to be accountable

and implement safety measures to ensure the safety of their users. The Minister for Communications, the Hon Michelle Rowland MP has expressed that 'by the sheer size, market dominance and influence, these platforms are also the site of a high information asymmetry and power imbalance. Many platforms have taken on some responsibility, establishing terms of service and content policies to address online harms but it's clearly not enough'. (Rowland MP, 2023)

There is a necessary balance that must be considered between the profits of industry and the indirect costs that result from the profit such as the harms users experience. This has been a consideration throughout the development of the standards and our policy discussion outlined in earlier questions. Ultimately, the profits of industry cannot supersede or take precedence over the significant harms to users. Accordingly, more weight has been placed on the harm to users throughout the analysis as supported by evidence to highlight the significance of the problem that the standards seek to resolve.

Several gaps have been identified related to the standards. The main gaps include not having an exact figure of how many RES and DIS there are that are accessible to end-users in Australia. Additionally, we do not have precise knowledge of the safety systems and technologies these services are already operating. To overcome these gaps, we would rely on the implementation of the reporting requirements in the standards to obtain information via specific and annual reports. eSafety also intends to engage closely with the providers to seek their views on the standards and any gaps they identify.

## 6.2. Analysis of options

Each of the three potential options was considered against the decision criteria to ascertain the option that best meets the objective and guiding principles, including the outcome of the cost-benefit analysis, the qualitative factors which cannot be monetised and consultation feedback.

### 6.2.1. Summary of results of analysis of Option 1 (maintain the status quo)

A retention of the status quo (Option 1) does not provide adequate protection for Australian end-users due to the lack of uniform protections across RES and DIS

regarding class 1A and 1B material. The systemic presence of harmful content such as child sexual abuse material on some RES and DIS highlights that under the status quo the existing systems, process and technologies across the RES and DIS sections are either non-existent or inadequate to address the problem.

Option 1 (retention of the status quo) does not involve any cost to providers, due to there being no requirements to introduce increased protective systems, technologies, or policies in respect of class 1A and class 1B material on their services. Option 1 is the least costly option as there is no compulsory cost to industry due to the lack of mandatory legal requirements. However, in line with the decision-making guidelines, Option 1 not only does not meet the policy objective, but it is also not the most effective or appropriate option.

When considering all the factors, chapter 4 shows that Option 1 (maintain the status quo) has no regulatory cost burden to businesses, individuals, and community organisations. However, while not all costs of harms could be quantified for each of the policy options and types of material or the proportion to which each option might reduce harm (prevent costs), Option 1 is assessed to have no impact on reduce the harms from class 1A and 1B material on RES and DIS.

As highlighted throughout this impact analysis under the status quo there are not sufficient protections to address the risks of class 1A and 1B material on RES and DIS. As demonstrated by the research into the current scale and scope of its presence on RES and DIS, Option 1 would allow the production, distribution and consumption of seriously harmful and illegal material such as child sexual abuse material and pro-terror material to flourish at the cost of significant damage to individuals and communities, with a consequential flow-through effect to the Australian economy which bears the largely unquantifiable cost of this damage.

Option 1 does not meet the policy objective to promote and improve online safety for Australians in respect of class 1A and class 1B material. Option 1 – the status quo or ‘do nothing’ option - is therefore not a viable option.

### 6.2.2. Summary of results of analysis of Option 2 (industry co-regulation)

Option 2 - registration of the draft RES and DIS codes - would provide some additional protections for Australian end-users from class 1A and class 1B material. However, these would not be sufficient to meet the policy objective. The

draft industry codes were not registered by the Commissioner as they failed to provide appropriate community safeguards, and do not provide adequate protections from class 1A and class 1B material. This was despite extensive consultation between eSafety and industry associations over an eighteen month plus period.

As discussed in chapter 4, Option 2 (industry co-regulation) has some regulatory costs to businesses in scope, although these are less than Option 3 - due to the draft codes having fewer obligations than the standards and the reduced enforceability of key obligations. While eSafety was not able to quantity the costs of all harms for each of the policy options and the proportion to which each option might reduce the risk of such content on RES and DIS (and the consequent harm), Option 2 is likely to have some reduction in harms.

However, reflecting the decision that the draft industry codes for RES and DIS were found not to provide appropriate community safeguards, Option 2 would fail to adequately address the risk and harms associated with the production, distribution, and consumption of seriously harmful and illegal material such as child sexual abuse material and pro-terror material on these services. This has the cost of significant damage to individuals and communities, with a consequential flow-through effect to the Australian economy which bears a largely unquantifiable cost associated with this damage.

Due to the gaps in the regulatory framework which would allow bad actors to exploit weaknesses and the resulting costs to individuals and communities with Option 2, it is not considered a viable option.

### 6.2.3. Summary of results of analysis of Option 3 (direct regulation)

Option 3 – registration of the standards – is the recommended option as it returns the highest net benefit and meets the policy objective.

As outlined in chapter 4, in summary Option 3 (direct regulation) has the most significant regulatory costs for businesses in scope. However, this is balanced by the benefits expected to accrue to individuals, communities, and the Australian economy – through an anticipated but unquantifiable lowering of costs of harms, due to the expected decrease of class 1A and class 1B material on RES and DIS once the standards are fully operational.

**Option 3 (direct regulation) is estimated to have the greatest annual net benefit** while a benefit-cost ratio cannot be quantified (due to the absence of data on the harm/cost mitigations for each policy option). It is assessed that the implementation of the standards (Option 3) will highly likely lead to a reduction in the risk and growth of class 1A and class 1B on RES and DIS services, which will have a direct reduction in harms. Option 3 will provide a cost benefit to individuals, community, and government through a reduction in harms and associated economic, health and social impacts. Mitigation of these harms and associated costs (both tangible and intangible) is why Option 3 is considered to provide the greatest annual net benefit of the policy options.

The standards lay down a set of mandatory compliance measures, legally binding for all RES and DIS which can be accessed from Australia, requiring providers to:

- take proactive steps to create and maintain a safe online environment
- empower end-users in Australia to manage access and exposure to class 1A and class 1B material
- strengthen transparency of, and accountability for, class 1A and class 1B material on their services.

The standards will be regulatory instruments, and the obligations can be directly enforced including through civil penalties. Once the standards are registered, if a company fails to comply with an industry standard, this can result in a civil penalty of up to $782,500 or other enforcement actions.

The proposed measures ensure the highest level of accountability by RES and DIS providers to undertake actions to reduce material which causes these serious forms of online harms. Option 3 - direct regulation by registration of the standards - allows the eSafety Commissioner to provide for adequate regulation to which protect Australian end-users against class 1A and class 1B material across RES and DIS. This is consistent with the objectives of the Act in section 3 and the eSafety Commissioner's statutory functions in section 27 of the Act.

Option 3 achieves the best balance between the risk of harm to Australian end-users, their community, and the Australian economy from class 1A and class 1B material on RES and DIS, and the business interests of RES and DIS providers.

While Option 3 places greater responsibilities and a higher cost burden on RES and DIS providers, the standards are risk-based, with requirements placed on providers proportionate to the risk their service presents in respect of class 1A and 1B material. The requirements in the standards are also outcomes-based, setting out the objectives while remaining technology neutral, and allowing providers to choose how best to meet the required outcomes within their existing framework of operations.

The standards also include amendments made to address concerns expressed by industry (and other stakeholders) during consultation. These changes help ensure the obligations are achievable, practical and flexible while ensuring the protections against highly harmful material to be put on place on their services are meaningful.

Option 3 – implementation of the standards - is the best option as it is the most effective, appropriate and efficient way to best meet the policy objective. The RES and DIS standards offer the highest net benefit and in accordance with the decision-making principles is our recommended option.

## 6.3. Implementation plan

The implementation of Option 3 (direct regulation) will require a coordinated effort between government bodies. The standards will come into effect 6 months after the day they are registered on the Federal Register of Legislation. A timeline below highlights both the implementation of the standards and the key reporting requirements under the standards for service providers.

The key implementation stages include:

- As a delegated instrument the final standards require registration under the Legislation Act 2003 with accompanying documents The standards will be tabled before parliament with their supporting documents including the Impact Analysis and Explanatory Statement.
- The standards and supporting documents will be published on the eSafety website.
- eSafety will develop and publish Regulatory Guidance on the standards.
- The standards and their compliance obligations come into effect 6 months after registration. Service providers would be required to adhere with their legislative obligations from this date.

- For those providers required to submit annual compliance reports under the standards, annual reports will commence 12 months after the commencement of the standards.

### 6.3.1. Implementation challenges and risks

Implementation of the standards has the following key challenges and risks:

- Providers not understanding the requirements of the standards
- Providers not agreeing with requirements and intentionally not complying with their new obligations.
- Some overseas-based service providers may maintain that they are not obligated to comply with Australian law.
- Balancing eSafety's regulatory role in a rapidly evolving online safety landscape, where technology and services are constantly changing.
- Broader regulatory developments in generative AI may ensure regulatory coherence is difficult to maintain.

To mitigate these low-level risks eSafety will develop regulatory guidelines to assist providers to understand and comply with their obligations. eSafety will continue to regularly engage with industry, and conduct ongoing stakeholder meetings, including with RES and DIS providers, to assist them with understanding the requirements, encourage compliance, and hear first-hand industry feedback and observations. Should further concerns arise with online safety risks in relation to the class 1A and class 1B material on emerging services such as generative AI services or the policy landscape in relation to generative AI services evolves in a particular way, eSafety can consider whether a further standard is required.

The requirements in the standards are proportionate to the broad risk associated with different types of services regarding class 1B material. Providers of categories of services with minimal to no risk will not be subject to the obligations under the standards (e.g. DIS falling within Tier 3). The standards also have exemptions such as technical feasibility and a test of 'appropriateness' for many of the measures avoiding the placement of unreasonable obligations on providers. The providers of some services may already have systems, processes and/or technologies in place to fulfill certain obligations, resulting in reduced initial cost burdens.

### 6.3.2 Transitional Arrangements

As outlined in 6.3.1 implementation of the two standards has several associated challenges and risks. A provider's ability to meet the requirements in the standards is also dependent on a variety of factors.

To facilitate the smooth introduction of the standards eSafety has prepared the following transitional arrangements:

- Upon tabling the standards in Parliament eSafety will publish a media release and relevant documents on our website to inform industry of the registration of the final standards. As the standards do not come into effect until 6 months after their registration, this transition period will give providers appropriate time to understand the new regulatory requirements, determine what compliance obligations are applicable to them and meet these requirements.
- During this 6-month period eSafety will provide support to industry to assist them with interpretation of the standards. In addition to outlining relevant policy intent, eSafety will publish regulatory guidance, fact sheets, Q and A documents and other information to help inform industry of the standards obligations.
- Annual reports will not commence until 12 months after the standards come into effect, giving providers an adequate period to obtain the necessary systems, processes, or technologies that their service(s) require to comply with the standards.
- eSafety will conduct regular engagement with RES, DIS providers and relevant stakeholders such as industry associations. While eSafety is unable to provide legal advice, industry can contact eSafety with queries.

# 7. How will you evaluate your chosen option against the success metrics?

**In this chapter we describe how we will evaluate the performance of the RES and DIS industry standards against the objectives and success measures outlined in Question 2, during and after implementation.**

## 7.1. The policy objective and the standards

As detailed in Chapter 2, the objectives of the RES and DIS industry standards (which are in section 4 of each standard) are to improve online safety for Australians in respect of class 1A material and class 1B material, including by ensuring that providers of RES and DIS services establish and implement systems, processes and technologies to manage effectively risks that Australians will solicit, generate, distribute, get access to or be exposed to class 1A material or class 1B material through the services.

## 7.2. Performance monitoring and evaluation

The objectives and success metrics set out in question 2 will require monitoring of providers' compliance with the standards to ensure their implementation and ongoing operation continues to meet the policy objectives.

Table A below provides a broad overview of the various measures eSafety will use to evaluate the standards against the success metrics.

### Table 10 - Measures eSafety will use to evaluate the standards against the success metrics

**Objective: improve online safety for Australians in respect of class 1A and class 1B material on RES and DIS services**

| Success measures | Evaluation metrics |
|---|---|
| RES and DIS providers engage with the standards | 1. Annual reports required under the RES industry standard are received by eSafety within the required timeframe.<br>2. Number[49] of risk assessments provided by Tier 1 RES to eSafety under annual report requirement.<br>3. 90% of the following notices issued to RES and DIS providers receive a response from the industry participant within the required timeframe:<br>a. Risk assessments and other information;<br>b. Reports of technical feasibility, systemic vulnerability etc of provisions of Division 2;<br>c. Outcomes of development programs; and<br>d. Compliance and other certifications and reports;<br>4. Number[50] of new features notified to eSafety. |
| Class 1A material is proactively detected and removed by RES and DIS providers | 5. The proportion of child sexual exploitation material and pro-terror material that providers have identified and acted against, as reported to eSafety under the RES standard annual compliance reporting, and f the DIS standard annual compliance reports. |
| Positive safety interventions have been taken by RES and DIS providers | 6. Across the reporting period eSafety will track the introduction of online safety interventions by RES and DIS providers which can be the standards have contributed to, such as introduction of user reporting options, through reports provided and responses to such BOSE notices as may be issued.<br>7. eSafety will track at a broad level the likely compliance cost incurred by RES and DIS providers which can be attributed to the standards to maintain or introduce positive safety interventions. This could be inferred through annual compliance reports, reports on outcomes of development programs, reports of technical feasibility. Other information such as responses to BOSE notices and publicly available information may also assist. |
| Feedback from stakeholders on the effectiveness of the RES and DIS industry standards | 8. Feedback from stakeholders as to whether they consider the standards are effective in increasing online safety in respect of class 1A and class 1B material across RES and DIS services. Stakeholder could include (but are not limited to) the National Centre for Missing and Exploited Children, Tech Against Terrorism, researchers, academics, and community safety advocates.<br>9. Feedback from providers on compliance costs incurred because of implementing and complying with the applicable standard (given the uncertainty of regulatory burden estimates). |

[49] A percentage metric is not possible for this measure, as we do not know the total number of Tier 1 RES.
[50] A percentage metric is not possible for this measure, as it is not possible to identify all new features and assess industry participants' compliance.

## 7.3. Complicating factors

Development of these metrics has been complicated as there is limited baseline data available against which to measure improvements directly caused by the RES and DIS standards:

- the exact number of RES and DIS providers is unclear and there is no exhaustive list of RES and DIS providers impacted by the standards.
- the proportion of RES and DIS providers who already have measures, technologies and systems in place is currently unknown – as is also the extent to which these measures are effective against risks from class 1A and class 1B material.

The success measures have therefore been designed around measuring industry engagement with the standards, and with metrics designed to allow establishment of a baseline for high-risk providers.

## 7.4. Ongoing evolution of the performance metrics

Following the registration of the standards, eSafety will develop a program to monitor compliance with the new enforceable obligations under the standards, including receiving, investigating, and monitoring complaints in relation to potential breaches of the standards. This will lead to investigations and enforcement action where necessary and will sit alongside eSafety's powers in relation to the registered industry code. The information obtained will also contribute to evaluating the effectiveness of the standards and will allow for the iterative evolution of the performance metrics.

If certain provisions of the standards prove ineffective against its intended outcomes, eSafety may consider varying the standards to ensure the risk that Australians will solicit, generate, distribute, get access to or be exposed to Class 1A and 1B material through a RES or DIS is effectively managed. Variation may also be necessary given the evolution of the generative AI ecosystem.

Section 148 of the Act includes a requirement for mandatory consultation for any variations to an industry standard that are not considered 'minor'. The Commissioner is required to make a copy of the draft available on the eSafety

website and invite interested persons to provide comments over a minimum 30-day period. Subsequently, due regard must be given to comments before varying the industry standard. This will provide a useful way to monitor the effectiveness of Option 3. As the regulatory and online ecosystem changes over time obtaining feedback from the public will ensure valuable contributions about the current standards and any proposed amendments to ensure its effectiveness as a regulatory instrument.

In combination, all the above measures and metrics will ensure that the effectiveness of the implementation of Option 3 (direct regulation) through the standards will continue to be actively monitored and evaluated against their objectives during and post the implementation period.

# 8. Glossary

A number of these terms are defined in either the OSA or the standards. Readers are advised to read section 6 of each of the standards for the full definition which will apply legally.

**AI image generators -** refers to the process of using machine learning to create visual content from text prompts, ranging from realistic images to illustrations.

**App/application -** an app is like a computer program but is designed to work on the small screen of a smartphone or tablet. Some apps don't need the internet to work, but many apps do.

**App distribution services -** means a service that enables end users to download apps, where the download of the apps is by means of a carriage service.

**Australian Centre to Counter Child Exploitation (ACCCE) -** is led by the Australian Federal Police and works with public and private sections, as well as civil society, to drive a collaborative national response to counter the exploitation of children in Australia. ACCCE focuses on countering online child sexual exploitation, and as such, organised child exploitation networks operating in the online environment.

**Australian Security Intelligence Organisation (ASIO) -** is Australia's national security agency responsible for the protection of the country and its citizens from espionage, sabotage, acts of foreign interference, politically motivated violence, attacks on the Australian defence system, and terrorism.

**Australian Federal Police (AFP) -** is the national and principal federal law enforcement agency of the Australian Government with the unique role of investigating crime and protecting the national security of the Commonwealth of Australia

**Child sexual abuse material (CSAM) -** means material that: (a) describes, depicts, promotes, or provides instruction in child sexual abuse; or (b) is known child sexual abuse material.

**Child sexual exploitation material (CSEM) -** means material that: (a) is or includes material that promotes, or provides instruction in, paedophile activity; or (b) is or includes: (i) child sexual abuse material; or Interpretation (ii) exploitative or offensive descriptions or depictions involving a person who is, appears to be or is described as a child; or (c) describes or depicts, in a way that is likely to cause offence to a reasonable adult, a person who is, appears to be or is described as a child (whether or not the person is engaged in sexual activity); and, in the case of a publication, also includes material that is or includes gratuitous, exploitative or offensive descriptions or depictions of: (d) sexualised nudity; or (e) sexual activity involving a person who is, appears to be or is described as a child.

**Classified -** means classified under the Classification (Publications, Films and Computer Games) Act 1995

**Class 1A material -** child sexual exploitation material, pro-terror material, and extreme crime and violence material

**Class 1B crime and violence material -** refers to material that describes, depicts, expresses, or otherwise deals with matters of crime, cruelty or violence without justification, and material that promotes, incites, or instructs in matters of crime or violence.

**Class 1B drug-related material -** refers to any material that describes, depicts, expresses, or otherwise deals with matters of drug misuse or addiction without justification, or which instructs or promotes drug use.

**Cloud computing -** is running programs and services over the internet on equipment owned by someone else. An example is an online service that allows you to upload and store photos online – in 'the cloud' – so you can access them as needed from a computer, smartphone, tablet, or other device.

**Deepfake -** A 'deepfake' is an extremely realistic – though fake – image or video that shows a real person doing or saying something that they did not actually do or say. Deepfakes are created using artificial intelligence software that draws on many photos or recordings of the person. Deepfakes have been used to create fake news, celebrity pornographic videos and malicious hoaxes.

**De-platforming** – refers to the barring of individuals, groups, or entities from sharing their views or content on a digital platform.

**Designated internet service -** is defined in section 14 of the OSA. It is a very broad category of services that includes online services not covered by the other industry section. It will include many apps and websites, as well as online storage services which are used by end-users to upload, store and manage their files including photos and other media. Examples include websites (excluding social media, email, chat, messaging, online gaming and dating sites), and consumer cloud storage such as iCloud, Dropbox, OneDrive and Google Drive.

**End-to-end encryption (E2EE) -** describes a means of securing communications from one device, 'sender', or 'end point', to another intended recipient. E2EE transforms standard text, imagery, and audio into an unreadable format while it is still on the sender's system or device.

**End-user -** is the person who uses a piece of software or an online service.

**End-user managed hosting services –** refer to file or image storage services.

**Extreme crime and violence material -** in relation to a computer game, means material that is crime and violence material in relation to a computer game where, without justification, the impact of the material is extreme because: (a) the material is more detailed; or (b) the material is realistic rather than stylised; or (c) the game is highly interactive; or (d) the gameplay links incentives or rewards to high impact elements of the game; or (e) for any other reason.

**File or image storage services -** are types of end user managed hosting services. Examples of end-user managed hosting services include online file storage services, photo storage services, and other online media hosting services, including such services that include functionality to allow end-users to post or share content.

**Generative Artificial Intelligence (Generative AI) -** refers to a branch of AI that develops generative models with the capability of learning to generate content such as images, text, and other media with similar properties as their training data.

**Gore sites -** serve as digital hubs for the sharing of real-life killings, torture, and other forms of violence, catering primarily to ‘gore seekers’; a niche audience searching for graphic and disturbing material.

**Grooming -** is when an adult deliberately establishes an emotional connection with a child to lower their inhibitions, to make it easier to have sexual contact with them. It may involve an adult posing as a child in an online game or on a social media site to befriend a child and encourage them to behave sexually online or to meet in person.

**Hash or hashing -** is a one-way cryptographic function that generates a summarised character string, known as a hash, from a data record. For example, a hash of an email address may be used to search in a database without sharing the content. A data record can be a word, a sentence, a longer text or an entire file.

**industry code -** has the meaning given in section 132 of Online Safety Act (2021).

**Image-based abuse -** refers to sharing, or threatening to share, an intimate image or video without the consent of the person shown.

**Immersive technologies -** enable a user to experience and interact in three-dimensions (3D) with digital content in a way that looks, sounds, and feels almost real. These technologies include augmented reality (AR), virtual reality (VR), mixed reality (MR) and haptics (interaction involving touch).

**Known child sexual abuse material -** means material that: (a) is or includes images (either still images or video images); and (b) has been verified as child sexual abuse material by a governmental (including multi-lateral) or non-governmental organisation: (i) the functions of which are or include combating child sexual abuse or child sexual exploitation; and (ii) in the case of a non-governmental organisation—that is generally recognised as expert or authoritative in that context; and (c) is recorded on a database that: (i) is managed by an organisation of a kind described in paragraph (b); and (ii) is made available to government agencies, enforcement authorities and providers of designated internet services for the purpose of their using technological means to detect or manage child sexual abuse material on designated internet services.

**Known pro-terror material -** means material that has been verified as pro-terror material. Note 1: Known pro-terror material may include material that can be detected via hashes, text signals, searches of key words terms, or URLs or behavioural signals or patterns, that signal or are associated with online materials produced by terrorist entities that are on the United Nations Security Council's Consolidated List.

**Live streaming -** refers to online media that is simultaneously recorded and broadcast in real time to the viewer. All you need to be able to live stream is an internet-enabled device, like a smartphone or tablet, and a platform (such as website or app) to broadcast on. Live streaming does not normally involve two-way audio and video communication, although may occur on services with these features.

**Large Language Model (LLM) -** refers to a type of artificial intelligence algorithm that uses deep learning techniques and large data sets to understand, summarise, generate and predict new content.

**Model Distribution Platform -** means a designated internet service with the predominant functionality of making available one or more machine learning models and making such models available for download.

**Machine learning -** is an approach or effort that uses algorithms to process expanding data sets so computing systems can further expand and refine the outputs. The data sets are effectively the experiences that the systems 'learn' from. As machine learning improves, the systems may give the impression of approaching 'artificial intelligence'.

**Moderator -** Some social media services and online chat rooms and forums assign moderators with special privileges to check and manage the content of conversations to ensure that users participate according to the site rules. Moderators are often able to block both individual comments and users who do not participate appropriately. They generally aim to keep conversations on topic in an unbiased manner in line with the forum's guidelines.

**Multimodal (AI) models -** is a technology that can handle and process a wide variety of inputs, including text, images, and audio, as prompts and convert those prompts into various outputs, not just the source type.

**National Center for Missing & Exploited Children (NCMEC) -** is a private, nonprofit organisation whose mission is to help find missing children, reduce child sexual exploitation, and prevent child victimisation. NCMEC operates a CyberTipline which processes and reviews reports of child sexual exploitation (including sexual abuse, online enticement, and contact offenses) and shares them with law enforcement agencies.

**Open source -** refers to publicly available information.

**Peer-to-peer (P2P) networking -** Peer-to-peer applications run on a personal computer or other digital device and share files, such as music or videos, with other online users. Peer-to-peer networks connect individual computers together to share files instead of having to go through a central server.

**Private messaging service -** is a type of communication wherein the message can only be viewed or read by a specific recipient or group of people.

**Pro-terror material -** includes any material that directly or indirectly counsels, promotes, encourages, instructs, or urges a terrorist act. Class 1A extreme crime and violence material includes content that shows, describes, promotes, incites, or instructs people in violent crimes including terrorist acts, kidnapping with violence or threats of violence, murder, attempted murder, rape, torture, and suicide.

**Provide/provider -** refers to a relevant electronic or designated internet service that makes the service available.

**Relevant electronic service -** is defined in section 13A of the OSA. It broadly refers to those online services that enable end -users to communicate with one another, including email, instant messaging, short message services, multimedia message service, online gaming and dating services.

**Safety by Design -** is an eSafety initiative that places the safety and rights of users at the centre of the design, development and deployment of online products and services. The initiative aims to assist industry to take a proactive and consistent approach to user safety when developing online products and services. It seeks to create stronger, healthier, and more positive communities online by driving-up standards of user safety.

**Service -** refers to a relevant electronic or designated internet service.

**Sexual extortion -** refers to someone who tries to blackmail a person over intimate images or videos of them. This is a type of image-based abuse called sexual extortion, sometimes known as sextortion. The blackmailer threatens to reveal intimate images of the person unless they give in to their demands. These demands are typically for money, cryptocurrency, gifts cards, online gaming credits or more intimate images.

**User-generated content -** is any form of content – such as a text, post, image, video or reviews – created by an individual (not a brand, company or organisation) and posted or shared online.

**Uniform Resource Locators (URL) -** URL stands for a 'uniform resource locator', such as an address of a file or webpage.

**Voice over internet protocol (VoIP) -** is a technology that allows voice to be transmitted using the same protocols – or sets of rules – that the internet uses. Skype, WhatsApp, and Facebook Messenger, for example, all use VoIP technology to allow users to make calls.

# 9. Annexures

## 9.1. Annexure A – Classification and categorisation of class 1A material

| Class 1 & 2 (Part 9 Online Safety Act) | Subcategories of material to be dealt with by codes | Online material | eSafety harms lens |
|---|---|---|---|
| Class 1 (RC) | 1A | **CSEM Child sexual exploitation material.** Material that promotes or provides instruction of paedophile activity.<br><br>**Pro-terror content** Material that advocates the doing of a terrorist act (including terrorist manifestos).<br><br>**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. | **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<br><br>**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<br><br>**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 |

# 9.2. Annexure B – Regulatory burden estimate assumptions, limitations, and methodology

## 9.2.1. Methodology notes

### 9.2.1.1. Estimating the number of businesses in scope

Several data sources were used to estimate the number of Australian businesses impacted by the policy options. These included specialised ABS research and publicly available data. Compound growth rates over a ten-year period were also calculated on each data set. These data sources and some identified assumptions and limitations in using these sources are outlined below.

It is highlighted that the assumptions identified here are not exhaustive and that there are almost certainly more assumptions and limitations that underpin the use of the below data sources. Given the inability to undertake cost benefit analysis this approach was considered to provide the most realistic assessment of estimated regulatory costs to Australian businesses.

## 9.2.2. Data sources used

### 9.2.2.1. Australian Bureau of Statistics (ABS) Australian and New Zealand Standard Industrial Classification (ANZSIC) data

Australian Bureau of Statistics (ABS) Australian and New Zealand Standard Industrial Classification (ANZSIC) data was used as a first source to estimate the number of RES and DIS in scope. Data was sourced for the following two ANZSIC codes.

- **5802** - Other Telecommunications Network Operation (used to indicate RES).
- **5700** – Internet Publishing and Broadcasting (used to indicate some websites - designated internet services).

Based on ANZSIC codes description and primary activities, these two codes provided the best representation of some critical RES and DIS services.

The number of businesses for these two codes are based on information from the ABS in the following data sets by businesses (employee size):

- 'Counts of Australian Businesses, including Entries and Exits June 2019 to June 2023', using 'Data Cube 2: Businesses by Main State by Industry Class by Employment Size Ranges'.[51] (Australian Bureau of Statistics, 2023) and
- 'Counts of Australian Businesses, including Entries and Exits June 2015 to June 2019', using 'Data Cube 2: Businesses by Main State by Industry Class by Employment Size Ranges'.[52] (Australian Bureau of Statistics, 2020) *This data covers the financial year 2018-2019.*

Non-employing businesses and (micro businesses) (0-4 employees) have been included in the impact analysis and are captured under small businesses, as the size of a business does not preclude them from undertaking activities that would be subject to compliance obligations under the policy options.

#### 9.2.2.2. Estimated 10-year compound growth 10-year outlook on the number of businesses impacted – ANZSIC data

To determine the compound growth of the number of businesses under the two ANZSIC codes, the data sets (above) which cover a five-year period were used to calculate the compound growth rate over this period. This rate was then applied to the current number to calculate the expected number of businesses in 10 years (compound growth rate). Table 1 below provides the data and rates for determining the compound growth.

**Table 1: Estimated ANZSIC data (5700 and 5802) compound business growth over 10 years**

| Count | ANZSIC Industry Code | ANZSIC Industry | Total all Businesses 2018-2019 | Total all Businesses 2022-2023 | Compound Annual Growth Rate (last 5 years observed) | Compound Annual Growth Rate (next 10 years calculated) 2033 |
|---|---|---|---|---|---|---|
| | Code | Description | no. | no. | no. | no. |
| Total count | 5802 | Other Telecommunications Network Operation | 522 | 682 | 5.492789012 | 1164 |
| Total count | *5700* | *Internet Publishing and Broadcasting* | 1,555 | 1,833 | 3.344272967 | 2547 |

[51] Counts of Australian Businesses, including Entries and Exits, July 2019 - June 2023 | Australian Bureau of Statistics (abs.gov.au)
[52] Counts of Australian Businesses, including Entries and Exits, July 2015 - June 2019 | Australian Bureau of Statistics (abs.gov.au)

| Count | ANZSIC Industry Code | ANZSIC Industry | Total all Businesses 2018-2019 | Total all Businesses 2022-2023 | Compound Annual Growth Rate (last 5 years observed) | Compound Annual Growth Rate (next 10 years calculated) 2033 |
|---|---|---|---|---|---|---|
| **Total Count** | | | **2,077** | **2,515** | | **3711** |

### Assumptions regarding ANZSIC data

ANZSIC data captures only some of the Australian businesses operating RES and DIS services in scope of the policy options and includes some services which will not have any obligations under the standards. A key limitation is that the data cannot be disaggregated to extrapolate a more accurate sample of in scope services, therefore it is almost certain that many of the services in these categories selected (based on description and primary activities) are not in scope of the policy options. For example, 5700 also captures social media services, and 5802 also captures other types of communications (for example satellite communications). These types of services are covered by under existing industry codes or otherwise not in scope of the RES and DIS standards ANZIC industry data also pertains only to registered Australian businesses and therefore RES and DIS providers that are not operating as a registered business may not be captured (websites etc).

#### 9.2.2.3. Digital Game Development Businesses

Online gaming services are RES and in scope of regulatory burden estimates. To determine and estimate the number of Australia gaming services, data was sourced from the ABS released data on Film, Television and Digital Games, Australia - Digital game development businesses.[53] This data provided the number of Australian registered digital game development businesses operating at the end of 2015-16 and 2021-22 financial year.

[53] Film, Television and Digital Games, Australia, 2021-22 financial year | Australian Bureau of Statistics (abs.gov.au)

**Table 2: ABS Digital Game Development Businesses 2015-2016 and 2021-2022**

**Film, Television and Digital Games, Released 22/06/2023**

| Location | Businesses at end June | Businesses at end June |
|---|---|---|
| | 2015-16 | 2021-22 |
| | no. | no. |
| Australia | 80 | 188 |

#### 9.2.2.4. Estimated 10-year compound growth on the number of businesses impacted – ABS Digital Game Development Businesses

To determine the compound growth of businesses for t 6659073his data, the data sets (above) which cover a six-year observed period (2015-2019 to 2021-2022) were used to calculate the compound growth rate over this period. This was then applied to the current rate to calculate the expected number of businesses in 10 years. Table 3 below provides the data and rates for determining the compound growth.

**Table 3: Estimated ABS Digital Game Development Businesses compound growth over 10 years**

| Count | ABS Data | Businesses at end June 2015-2016 | Businesses at end June 2021-2022 | Compound Annual Growth Rate (last 6 years observed) | Compound Annual Growth Rate (next 10 years calculated) 2032 |
|---|---|---|---|---|---|
| | | no. | no. | no. | no. |
| Total count | Digital Game Development Businesses | 80 | 188 | 15.30407171 | 781 |

**Assumptions on ABS Digital Game Development Business Data**

According to ABS methodology for the data, 'businesses were also coded as Digital game development businesses based on detailed financial data reported in the collection'. As there is no unique ANZSIC category for digital game development services, a list of digital game development businesses was initially manually compiled by the ABS. Adjustments were then made to remove the contributions of businesses that were found to be incorrectly coded as Digital game development businesses.

Not all these game development businesses captured will necessarily have communications functionalities, but it is expected that many will, and this data likely provides the most accurate estimate of the number of Australian online gaming businesses. There are likely to be variables which will impact on the growth of digital game development and historical growth may not represent future growth.

#### 9.2.2.5. Australian Dating Services

Dating services are a RES, however the dating services most used by Australian end-users are global businesses. To determine the number of Australian dating sites, data was sourced from the Australian Competition and Consumer Commission (ACCC) Online dating industry report 2015.[54] This data was estimated by the ACCC who swept dating site domains to determine the number that were Australian based.

**Table 4: ACCC Australian Online Dating Sites – 2014-2015**

| Count | Category | Total Australian domains 2014-2015 |
|---|---|---|
| Total count | Online Dating Sites (Australia) | no. |
| | | 31 |

#### 9.2.2.6. Estimated 10-year compound growth on the number of businesses impacted – Australian Online Dating Sites

Compound growth rate could not be determined from this data as no observed measurements were available. Estimated industry growth rate was obtained from a secondary data source and applied to the primary data. Industry growth rate was obtained from IBISWorld data for Dating Services in Australia 2024-2029. This source estimated that there had been a growth rate in the industry/number of businesses between 2019 and 2024 of 7.7 percent.[55] This growth rate was applied to the data obtained from the ACCC report to determine a growth figure and likely number of dating services at the end of 10 years.

[54] Online dating industry report (accc.gov.au) Online dating industry report (accc.gov.au)
[55] Dating Services in Australia - Market Size, Industry Analysis, Trends and Forecasts (2024-2029)| IBISWorld

**Table 5: Estimated Australian Dating Sites compound growth over 10 years from 2024**

| Count | Industry | Total number of Australian dating domains 2014-2015 (ACCC) | Industry Growth Rate 2019-2024 (Five Years) | Compound Annual Growth Rate | Compound Annual Growth Rate (next 10 years calculated) 2024 | Compound Annual Growth Rate (next 10 years calculated) 2034 |
|---|---|---|---|---|---|---|
| Total count | Online Dating Sites (Australia) | no. | % | | no. | no. |
| | | 31 | 7.70% | 1.864638 | 37 | 45 |

### Assumptions on Australian Dating Services Data

The key assumption is that the growth rate sourced from IBISWorld is an accurate reflection of the industry, because it has been sourced from a different data set. A limitation in this data was the absence of a repeat study that could enable the determination of growth on the same source and methodology. The methodology or assumptions and data sources used by IBISWorld to determine their growth rate for Australian dating services was not available.

The resulting figures in Table 5 (37 in 2024 and 45 in 2034) are expected to significantly overestimate the number of Australian dating services. As at the date of the preparation of this assessment, eSafety is only aware of five Australian dating services.

#### 9.2.2.7. Australian Based App Developers

There were limited data sources available to determine the number of Australian based app/application developers who are DIS and in scope of compliance obligations in the DIS Standard (or draft DIS codes). Data was sourced from Google Play for the number of Australian-based developers on its service in 2024.[56]

[56] Supporting the Thriving and Competitive Mobile Ecosystem in Australia (blog.google)

**Table 6: Total Number of Australia-based developers on Google Play**

| Year | Total Number of Australia-based developers on Google Play |
|---|---|
| 2024 | 12,200 |

#### 9.2.2.8. Estimated 10-year compound growth on the number of businesses impacted – Australia-based developers

Compound growth rate could not be determined from this data alone because only one observed measurement was available. Estimated industry growth rate was obtained from a secondary data source and applied to the primary data (Google). This source estimated that there had been a growth rate in the mobile application market in Australia of 7.7 % between 2022-2026.[57] This growth rate was applied to determine the estimate of growth to the data obtained from Google.

**Table 7: Estimated Australia-based developers compound growth over 10 years**

| Date | Total Number of Australia-based developers on Google Play | Growth rate over 4 years (2022-2026) | Compound Annual Growth Rate (over 4 years observed) | Estimated Total Australia-based developers on Google Play – with annual growth (next 10 years calculated) 2033 |
|---|---|---|---|---|
| 2024 | 12,200 | 0.077 | 1.871787318 | 14,686 |

#### Assumptions – Australian App Developers

The key assumption is that the growth rate sourced from Statista and applied to the primary data is an accurate reflection of the industry growth. The methodology or assumptions and data sources used by Statista to determine their growth rate for Australian dating services was not available and may not be comparable.

This original figure to which the growth rate is applied also only represents one data source, and figures from other key app stores, such as Apple, were not available. Therefore, this may underrepresent the number of Australian app developers (although most of the Australian app developers developing apps for

[57] Australia App Developers (2024) - Business of Apps; https://www.statista.com/outlook/dmo/app/australia

Google Play are also expected to make their apps available on Apple to ensure sufficient take up).

There are also likely to be variables which will impact on the growth of Australian app development in the next ten years and historical growth may not represent future growth.

#### 9.2.2.9. Australian Websites (Domains)

To capture DIS such as Australian websites, data was sourced from auDA (Australia's domain register) on the total number of Australian registered domains. The number of Australian domains was obtained for two financial years to estimate the compound annual growth.

**Table 8: Total Number of Australian Domains 2016-2017 and 2022-2023**

| Financial year | Total Australian domains under management |
|---|---|
| 2022-23 | 4,138,919 |
| 2016–17 | 3,111,507 |

#### 9.2.2.10. Estimated 10-year compound growth on the number of businesses covered – Australian Websites

To determine the compound growth of businesses for this data, the data sets (above) which cover a six-year observed period (2016-2017 to 2022-2023) were used to calculate the compound growth rate over this period. This was then applied to the current rate to calculate the expected number of businesses in 10 years. Table 9 below provides the data and rates for determining the compound growth.

**Table 9: Estimated Australia-based developers compound growth over 10 years**

| Financial year | Total Australian domains under management | Compound Annual Growth Rate (over 6 years observed) | Estimated Total domains – with annual growth (next 10 years calculated) 2033 |
|---|---|---|---|
| 2022-23 | 4,138,919 | 4.870 | 6,659074 |
| 2016–17 | 3,111,507 | | |

**Assumptions – Australian Domains**

Some Australian businesses operating websites will also use ".com" and potentially other domains so this data does not capture all Australian websites.

There are likely to be variables which will impact on the growth of websites over the next 10 years and historical growth may not represent future growth.

#### 9.2.2.11. Total RES and DIS estimated in scope of policy options (growth next 10 years calculated).

Table 10 below shows the consolidated data sources to estimates the number of RES and DIS services in scope of the policy options.

**Table 10 – Method 1 - Total RES and DIS estimated in scope of policy options (growth next 10 years calculated)**

| Data Source | Estimated No. with Compound Annual Growth (next 10 years calculated) |
|---|---|
| ANZSIC Code 5802 - Other Telecommunications Network Operation | 1164 |
| ANZIC Code 5700 - Internet Publishing and Broadcasting | 2547 |
| Digital Game Development Businesses | 781 |
| Online Dating Sites (Australia) | 45 |
| Australian registered domains (websites) | 6,659,074 |
| Australia-based developers (on Google Play) | 14,686 |
| **Total Number of Estimated Businesses/Services (in scope)** ***(rounded nearest hundred)*** | **6,700,000** |

#### 9.2.2.12. Total RES and DIS estimated in scope of regulatory burden costs (growth next 10 years calculated)

Australian websites and application developers are in scope of the DIS standard and have therefore been included in the overall estimate of businesses in scope (Table 10). However, most providers of websites, application developers and online services under ANZSIC code 5700 would have limited - to no – obligations under the DIS standard and therefore no regulatory costs.

There will be some specific Australian websites and apps that meet criteria set out in the standard which will be subject to meaningful obligations and costs, however there is no data that could be leveraged to measure what proportion of all Australian websites and apps that this subset would comprise.

It is estimated that there would not be many high-impact websites based on the UK OSA analysis, which identified only 11 'dedicated pornography providers' that were UK based platforms..[58]

It is estimated at a low range there would be 11 online services (given Australia's comparability with the UK environment) and a maximum of 100 online services as high impact online services. The median/average between these two (n=55) was selected as an estimate to represent high impact sites. High impact services are the key category of DIS with material obligations under the DIS Standard and so it is this figure which has been used to calculate the aggregate number in Table 11.

Regarding other categories of DIS with specific obligations under the DIS Standard, eSafety notes the following:

- **End-user managed hosting services (file and photo storage services)**
    - The end-user managed hosting services most widely used in Australia are not based in Australia.
- **High impact generative AI DIS**
    - eSafety understands that 'not safe for work' or specialised AI pornography generators, which would be captured by the high impact generative AI DIS category, are typically based overseas. The vast bulk of AI foundation models are made and operated by companies overseas ( (CSIRO, 2024)
- **Model Distribution Platform**
    - This is a small category of services, and eSafety is not aware of any based in Australia.
- **Tier 2**

[58] UK OSA Impact analysis page 42

- It is likely that there will be Tier 2 designated internet services based in Australia, however as Tier 2 services have less onerous obligations these have not been quantified.

Table 11 provides the estimate of the baseline of businesses in scope of regulatory cost burdens under the standards or equivalent codes. With the removal of ANZSIC code data 5700 (websites and application developers), **the number of RES and DIS in scope of the policy options is estimated to be 2045 (Table 11).**

### Table 11 – Method 1 - Total RES and DIS estimated in scope of policy options (growth next 10 years calculated)

| Data Source | Estimated Compound Annual Growth Rate (next 10 years calculated) |
|---|---|
| ANZSIC Code 5802 - Other Telecommunications Network Operation | 1164 |
| Digital Game Development Businesses | 781 |
| Online Dating Sites (Australia) | 45 |
| Australian registered domains (websites) * *Sample of all domains to represent estimate of high-risk internet sites* | 55 |
| **Total Number of Estimated Businesses/Services (in scope)** ***(rounded nearest hundred)*** | **2,045** |

### Assumptions

Key assumptions have been provided for each data set used to establish the baseline have already been canvassed.

There are two critical points to the methodology. The methodology's aim was to determine the number of Australian services with obligations under the policy options. Global businesses whose services are accessible to end-users in Australia are not covered. The data is also representative of those RES and DIS with obligations under the policy options. It is almost certain that this does not capture the wide range of services in scope, most of which will not have obligations under the standards (or the draft codes).

#### 9.2.2.13. Regulatory burden estimates

**Assumptions– Option 1**

Nil regulatory impact

**Assumptions Option 2 and 3**

Assumes there is nil regulatory cost impact on individuals or community organisations, as the regulation and associated costs are expected to only impact RES and DIS service providers which are businesses (this includes businesses that have no employees). While many community organisations will be DIS, they are expected to be Tier 3 and therefore not have any significant obligations under the DIS standard (or draft DIS code).

A key assumption in using the benefit transfer method is the reliance on and applicability of the secondary source data (in this case the Impact Analysis of the comprehensive UK OSA) is that the values are comparable (i.e., in location, scope, and other specific characteristics). The UK OSA impact analysis was selected as it had the closest comparability in terms of some of the services in scope and compliance requirements.

However, many of the services in scope of the UK OSA (including for example 'user-to-user' services (U2U services)[59] vary from the services in scope of policy options (ie those with obligations under the RES and DIS Standard). For example, the UK OSA applies to social media services, which are not in in scope of the Option 3. This impacts the comparability of the content moderation costs with Options 2 and 3. Social media services are likely to incur relatively significant costs under the UK legislation (due to volume of material) and therefore this is likely to significantly overestimate the relevant costs likely to be incurred by RES and DIS and, Australian *RES and DIS*, which would not have the same content moderation requirements. Other services such as email services, SMS and MMS are excluded from the UK OSA[60].

A further limitation is that there are significant differences between the obligations on services between the UK OSA and the policy options. In relation to RES, this is not expected to have an impact on overall cost estimates because

[59] 'User-to-user' services (U2U services) social media services; video-sharing services; messaging services; marketplaces and listing services; dating services; review services; gaming services; file sharing services; Search services, and Services that publish or display certain pornographic content.

[60]. Email, SMS (short messaging service), MMS (multimedia messaging service) and one-to-one live aural communications services are exempt services under the UK OSA.

most of these types of services (e.g. messaging services) are international, and except for dating and gaming services, are not costed for this impact assessment.

Importantly, the UK OSA captures a broader range of harms than the policy options considered here (the standards and the draft codes) which are limited to content which would likely be refused classification, if classified by the Classification Board. The UK OSA looks at a much broader range of harms including harms from:

- fraudulent advertising (scams),
- cyberbullying,
- cyberstalking
- online pornography (in terms of the impact on children)
- not protecting content of democratic importance etc.

The UK OSA also has broader obligations for transparency reports and risk assessment comparative to Option 2 and 3. Further, the UK costs include the impact from primary (comprehensive) legislation, related secondary legislation, and future codes of practice. Therefore, the scope of costs under the UK OSA would be an overestimate. Option 2 and 3 would be comparable to costs only associated with future codes of practice envisaged in the UK legislation. However, the proportion of costs related to only the future codes of practice could not be determined from the UK OSA impact analysis as the data was not further disaggregated.

As highlighted in text, there are several obligations under Option 2 and 3 that are not costed in the regulatory burden above, due to the absence of available data to obtain these estimates (i.e., these were not obligations under the UK OSA). Some of these provisions in the standards and codes include requirements for safety features and settings, resourcing trust, and safety, and ensuring that eSafety information is available to end-users. For Option 3 this also includes obligations for a program of investment and development activities (development program) in respect of systems, processes, and technologies.[61] These obligations will also require some regulatory costs to *applicable services but* have not been costed.

[61] Only applies to some high risk RES and DIS with monthly active end users over 1,000,000 in the previous calendar year.

In application of the benefit transfer methodology, extrapolation beyond the range of characteristics of the initial study is not recommended, however, extrapolation in future costs was required to estimate the regulatory burden over 10 years in line with impact analysis framework.

Benefit transfers can also only be as accurate as the initial value estimates and the veracity of data, and analysis that underpinned them. The UK OSA impact analysis involved a significant amount of research with UK industry to establish estimated costs and was based on statistically sound methodologies which aligned with UK regulatory and impact analysis framework. Analysis of the Australian impact analysis framework and the UK showed considerable comparability in requirements and policy considerations.

A further limitation of the methodology is that the unit value estimates can rapidly become dated. To compensate for this limitation the UK figures were adjusted for inflation from 2019 (source data) to 2023 rates. This was undertaken using the Reserve Bank of Australia online inflation calculator.[62] It is assumed that the inflation rates in Australia would be comparable to the UK over the period.

Assumptions on annual cost estimates: due to lack of available data it was not possible to disaggregate the estimated annual costs for startup costs versus ongoing costs to businesses, or between capital and labour costs. This inability to identify more disaggregated and specific costs to individual services is a limitation and data gap in estimating regulatory burden costs. Provision of per business cost assumes that each RES and DIS have equal obligations and regulatory burden costs. This is not expected to be the case.

As qualified within the impact analysis, each RES and DIS provider that is in scope of Option 3 will have a different regulatory burden. The extent of which will be determined by their risk classification, any existing mitigations (i.e., if they already have or are already undertaking requirements of the obligations and will not have implementation costs), technical feasibility or other limitations, and the size/turnover/complexity of the RES and DIS business). It is clarified that this represents an 'average' cost based on the total estimated regulatory burden for all providers estimated to have obligations under the policy options.

Assumptions projection of costs (10 years): There is a high degree of uncertainty in accurately projecting long term regulatory costs (i.e., 10-year projections), given

[62] Inflation Calculator | RBA

that both online safety technology as well as the services in scope of the Option 2 and 3 are rapidly evolving and developing (e.g, generative AI platforms and services).

It is highlighted that these assumptions are not exhaustive and that there are almost certainly more limitations that underpin the methodology used and the comparability of data sources. Given the inability to undertake cost benefit analysis this was determined to provide the most accurate assessment of estimated regulatory costs.

Assumption on cost variance between Option 2 and 3: It is assumed that the costs for Option 2 would not be as high as the costs for Option 3 on businesses in scope. This is because the draft codes (Option 2) do not have the same level of obligations as Option 3 and are also considered likely to be less enforceable. The compliance obligations for Option 2 and 3 were assessed against the UK OSA's assessment and assigned a comparable proportion of the UK cost estimates (see Table 12 and 13 of this annexure). This is an estimate only, as the exact cost burden variation cannot be precisely determined.

#### 9.2.2.14. Methodology for Regulatory Burden Measures

The following steps were undertaken to establish and transfer the compliance costs from the UK OSA Impact Analysis to Option 2 and 3:

- Transferable compliance areas and associated costs were extracted from the UK OSA Impact Analysis 'Summary of Impact' table[63] and then compared to the obligations of Option 2 – draft RES and DIS codes and Option 3- Draft Standards.
- The UK costs were adjusted for inflation (2019 to March 2024) and currency conversion to AUD (rates as at 13/05/2024).
- The UK costs were applied to the number of businesses in scope (n=2045) providing the relevant costs for each compliance obligation or comparable impact area.
- For Option 2 and Option 3 – the compliance obligations were compared to those UK OSA obligations through a qualitative assessment. A proportion was assigned to each compliance obligation (based on substantive

[63] United Kingdom Online Safety Bill Impact Assessment: https://assets.publishing.service.gov.uk/media/6231dc9be90e070ed8233a60/Online_Safety_Bill_impact_assessment.pdf pp 25-26, cited 15 May 2024.

requirements themselves in addition to enforceability and scope) for Option 2 and Option 3 and costs were then adjusted based on this estimate.

- The regulatory burden costs for Option 2 and Option 3 were estimated as relative proportions of the UK OSA estimates. The variation in compliance obligations was determined via a qualitative assessment of the draft industry codes (based on substantive requirements themselves in addition to enforceability and scope) and the UK OSA comparative obligations. A proportion was assigned to each compliance obligation for Option 2 and Option 3 and the costs were then adjusted based on this estimate.
- For example: re the UK OSA obligations '*undertaking additional content moderation*' it is estimated that for Option 2 (drafted codes), only 50% of the UK estimate should be apportioned and for Option 3 (Standards), approximately 80% of the UK estimate should be apportioned. This is because the UK OSA obligations for content moderation cover a much broader range of harms, not just illegal content, but also legal content which is harmful to end users.
- Estimates for Option 2 (drafted codes) reflect a lower proportion of the estimates associated with the UK obligations because the drafting of the obligations in the codes are less enforceable and requirements on service providers are not as extensive (they not provide the same level of safeguards) comparative to Option 3 (Standards).
- Costs are tabled to show the individual compliance obligation costs, total costs over 10 years, costs per business and annual costs to business.

These costs are provided in the following Tables (Table 12 – Option 2 and Table 13 – Option 3).

Table 12 – Option 2 Regulatory Burden Estimates by each compliance requirement -impact on RES and DIS in scope - costs over 10 years

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Code | Summary of Comparability: Comparability to Draft DIS Code | Estimated proportion of UK OSA costs: (0.00%) | Est costs for RES + DIS providers over a 10 Year Period (n =2045 businesses): $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (based on proportion) : $ (AUD) million | UK total estimates adjusted currency exchange (13-05-2024): $ (AUD) million | UK total estimates adjusted for inflation March 2024 (n=25000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019 prices) (n=25000): Low Estimate (£ million) |
|---|---|---|---|---|---|---|---|---|---|
| **Reading and understand the regulations** | In-scope platforms will be expected to familiarise themselves with the regulations which includes understanding which aspects of the safety duties apply to them and what steps they must take to ensure compliance. | Legal costs will be incurred to interpret and understand compliance obligations and information processing and dissemination. Transition costs. | Legal costs will be incurred to interpret and understand compliance obligations and information processing and dissemination. Transition costs. | **0.40** | **$1** | $1.92 | $23.45 | GBP 12.35 | GBP 9.60 |
| **Ensuring users can report harm** | Platforms will be expected to accommodate user reporting of harm and provide an avenue for user redress (challenge of content removal). User reporting and redress mechanisms are expected to vary across platforms. | MCM 19 - 'High risk' RES are required to have a reporting or complaints mechanism relative to whether the service can review and assess materials. For services that can assess and review materials, the RES Standard extends this provision by requiring the mechanism to be 'in service'. | MCM 23 - Tier 1, Tier 2 DIS and end-user managed hosting services required to have a reporting and complaints mechanism. | **0.80** | **$3** | $3.45 | $42.17 | GBP 22.21 | GBP 17.70 |

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Code | Summary of Comparability: Comparability to Draft DIS Code | Estimated proportion of UK OSA costs: (0.00%) | Est costs for RES + DIS providers over a 10 Year Period (based on proportion) : $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045 businesses): $ (AUD) million | UK total estimates adjusted currency exchange (13-05-2024): $ (AUD) million | UK total estimates adjusted for inflation March 2024 (n=25000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019 prices) (n=25000): Low Estimate (£ million) |
|---|---|---|---|---|---|---|---|---|---|
| **Updating terms of service** | All companies will be required to set terms of service for illegal content and, if relevant, protecting children. In addition, organisations will be required to set terms of service in relation to legal but harmful content | MCM 22 - 'High Risk/Tier 2' RES are required to publish and clearly communicate terms and conditions, community standards, and/or acceptable use policies broadly covering class 1A/B material. | MCMs 1, 15, 32 - Requirement for terms of service, community standards and/or other policies against enterprise customers being used to distribute illegal material, the storage of CSEM or pro-terror material on end-user managed hosting services, and class 1A material on Tier 1 and 2 DIS and end-user managed hosting services. | **0.50** | **$ 2** | $3.45 | $42.17 | GBP 22.21 | GBP 17.80 |
| **Conducting risk assessments** | All platforms in scope will be required to produce a risk assessment. Platforms will be expected to assess risks corresponding to the type of content and activity a platform is required to address | Clause 5 - Requirements for initial risk assessments if not a 'pre-assessed risk'. Requires risk assessment if there is a material change to the service. | Clause 4 - Requirements for initial risk assessments if not a pre-assessed risk. Requires risk assessment if there is a material change to the service. (Same as DIS Standard) | **0.70** | **$2** | $3.45 | 42.17 | GBP 22.21 | GBP 17.50 |

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Code | Summary of Comparability: Comparability to Draft DIS Code | Estimated proportion of UK OSA costs: (0.00%) | Est costs for RES + DIS providers over a 10 Year Period (based on proportion) : $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045 businesses): $ (AUD) million | UK total estimates adjusted currency exchange (13-05-2024): $ (AUD) million | UK total estimates adjusted for inflation March 2024 (n=25000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019 prices) (n=25000): Low Estimate (£ million) |
|---|---|---|---|---|---|---|---|---|---|
| **Undertaking additional content moderation** | Requirements for in scope platforms to put in place systems and process to address illegal content. Involve hiring additional content moderators, employing automated content moderations systems or a combination of both. **(Includes illegal and legal -but harmful)** | MCMs 3, 8-12. High risk RES that is capable of reviewing and assessing material on the service and removing material from the service will implement systems, processes, and/or technologies. These are unspecified and technology neutral. | MCM 8 requires DIS have systems, processes and/or technologies to detect and remove known CSAM. Applies to Tier 1 DIS. MCM 9 and 14 requires that DIS make ongoing investment in systems, processes and/or technologies which aim to disrupt and deter CSAM and pro-terror material, and tools and personnel to detect and remove class 1B material. Applies to Tier 1 DIS. MCM 6 requires DIS be reasonably resourced with personnel to ensure the safety of the service and operationalise the requirements of the Code. Applies to Tier 1, Tier 2 DIS and end-user managed hosting services | **0.50** | **$126** | $252 | $3,089 | GBP 1,627 | GBP 1,319 |

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Code | Summary of Comparability: Comparability to Draft DIS Code | Estimated proportion of UK OSA costs: (0.00%) | Est costs for RES + DIS providers over a 10 Year Period (based on proportion) : $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045 businesses): $ (AUD) million | UK total estimates adjusted currency exchange (13-05-2024): $ (AUD) million | UK total estimates adjusted for inflation March 2024 (n=25000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019 prices) (n=25000): Low Estimate (£ million) |
|---|---|---|---|---|---|---|---|---|---|
| **User verification and empowerment duties** | Platforms to offer optional user verification and provide user empowerment tools. In terms of optional user verification. Services would be required to put in place a mechanism by which an adult user could verify their identity. Separate from age assurance. | MCM 6 - most types of RES services to obtain use and retain registration details and have in place minimum user empowerment tools (e.g., blocking, age estimation technology, phone number, default private accounts, etc). | Not applicable to DIS | **0.80** | **$1** | $1.71 | $20.89 | GBP 11 | GBP 8.80 |

| Estimate | Est costs for RES + DIS providers over a 10 Year Period (based on proportion): $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045 businesses): $ (AUD) million | UK total estimates adjusted currency exchange (13-05-2024): $ (AUD) million | UK total estimates adjusted for inflation March 2024 (n=25000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019 prices) (n=25000): Low Estimate (£ million) |
|---|---|---|---|---|---|
| **Estimated Total Costs (all businesses over 10 year) Compound Annual Growth Rate (n= 2045)** | **$135** | $267 | $3,260 | | |
| **Estimated Costs (per business over 10-year Period) Compound Annual Growth Rate (n=2045)** | **0.07** | $0.13 | $1.59 | | |
| **Estimated Costs (annually) Compound Annual Growth Rate (n=2045)** | **$14** | $27 | $326 | | |

| | |
|---|---|
| *UK OSA in scope business* | 25000 |
| *In scope Australian business* | 2045 |
| | 8.18 |
| *%Difference* | 0.0818 |

Table 13 – Option 3 Regulatory Burden Estimates by each compliance requirement -impact on RES and DIS in scope - costs over 10 years

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Standard | Summary of Comparability: Comparability to Draft DIS Standard | Estimated proportion of UK OSA costs: (0.00%) | Est proportion of costs for RES + DIS providers over a 10 Year Period: $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045): $ (AUD) million | UK total estimates adjusted currency exchange (2024) (n=25,000) : Low Estimate (£ million) | UK total estimates adjusted for inflation March 2024 (n=25,000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019) (n=25,000): UK OSA Impact |
|---|---|---|---|---|---|---|---|---|---|
| **Reading and understand regulations** | In-scope platforms will be expected to familiarise themselves with the regulations which includes understanding which aspects of the safety duties apply to them and what steps they must take to ensure compliance. | Legal costs incurred to interpret and understand compliance obligations, information processing and dissemination across the business. Transition costs. | Legal costs incurred to interpret and understand compliance obligations, information processing and dissemination across the business. Transition costs. | **0.50** | **$0.96** | $ 1.92 | $ 23.45 | GBP 12.35 | GBP 9.60 |
| **Ensuring users can report harm** | Platforms will be expected to accommodate user reporting of harm and provide an avenue for user redress (challenge of content removal). User reporting and redress mechanisms are expected to vary across platforms. | Mechanisms to enable user reporting and complaints re material in breach of terms of use; standards complaints | Mechanisms to enable user reporting and complaints re material in breach of terms of use; standards complaints | **0.80** | **$2.76** | $3.45 | $42.17 | GBP 22.21 | GBP 17.70 |

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Standard | Summary of Comparability: Comparability to Draft DIS Standard | Estimated proportion of UK OSA costs: (0.00%) | Est proportion of costs for RES + DIS providers over a 10 Year Period: $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045): $ (AUD) million | UK total estimates adjusted currency exchange (2024) (n=25,000) : Low Estimate (£ million) | UK total estimates adjusted for inflation March 2024 (n=25,000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019) (n=25,000): UK OSA Impact |
|---|---|---|---|---|---|---|---|---|---|
| **Updating terms of service** | All companies will be required to set terms of service for illegal content and, if relevant, protecting children. In addition, organisations will be required to set terms of service in relation to legal but harmful content | Providers' terms of use must regulate use of the service and include obligations on account holders to ensure service is not used to distribute class 1A or 1B material and enable service provider to enforce terms of use. | Providers' terms of use must regulate use of the service and include obligations on account holders to ensure service is not used to distribute class 1A or 1B material and enable service provider to enforce terms of use. Applies to all DIS categories except Tier 3. | **0.60** | **$2.07** | $3.45 | $42.17 | GBP 22.21 | GBP 17.80 |
| **Conducting risk assessments** | All platforms in scope will be required to produce a risk assessment. Platforms will be expected to assess risks corresponding to the type of content and activity a platform is required to address | Providers which do not fall in pre-identified categories to produce a risk assessment on request. Risk assessment also required if a material change to the service. | Providers which do not fall in pre-identified categories to produce a risk assessment on request. Risk assessment also required if a material change to the service. | **0.80** | **$2.76** | $3.45 | $42.17 | GBP 22.21 | GBP 17.50 |
| **Undertaking additional content moderation** | Requirements for in scope platforms to put in place systems and process to address illegal content. Involve hiring additional content moderators, employing automated content moderations systems or a combination of both. **(Includes illegal and legal -but harmful)** | Detection and removal of known CSAM or pro-terror material for certain identified services subject to exceptions. Requirement is only for some RES categories reflecting risk associated with that type of service | Detection and removal of known CSAM or Pro-terror material from certain DIS services subject to exceptions. Requirement is only for some DIS categories reflecting risk associated with that type of service. | **0.80** | **$202** | $252 | $ 3,089 | GBP 1,627 | GBP 1,319.10 |

| Impact: UK OSA Impact | Summary of Comparability: What these cover (UK OSA Act) | Summary of Comparability: Comparability to Draft RES Standard | Summary of Comparability: Comparability to Draft DIS Standard | Estimated proportion of UK OSA costs: (0.00%) | Est proportion of costs for RES + DIS providers over a 10 Year Period: $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045): $ (AUD) million | UK total estimates adjusted currency exchange (2024) (n=25,000) : Low Estimate (£ million) | UK total estimates adjusted for inflation March 2024 (n=25,000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019) (n=25,000): UK OSA Impact |
|---|---|---|---|---|---|---|---|---|---|
| **User verification empowerment duties** | Platforms to offer optional user verification and provide user empowerment tools. In terms of optional user verification. Services would be required to put in place a mechanism by which an adult user could verify their identity. Separate from age assurance. | Obligations on providers to put in place safety features and settings to empower users as well as other obligations re provision of online safety information | Obligations on providers re safety features and settings but at a broader level than the RES user empowerment provisions. | **0.80** | **$1.37** | $ 1.71 | $ 20.89 | GBP 11 | GBP 8.80 |

| Estimate | Est proportion of costs for RES + DIS providers over a 10 Year Period: $ (AUD) million | Est costs for RES + DIS providers over a 10 Year Period (n =2045): $ (AUD) million | UK total estimates adjusted currency exchange (2024) (n=25,000): Low Estimate (£ million) | UK total estimates adjusted for inflation March 2024 (n=25,000): Low Estimate (£ million) | UK total estimates for all UK businesses over 10-year appraisal period (2019) (n=25,000): UK OSA Impact |
|---|---|---|---|---|---|
| **Estimated Total Costs (all businesses over 10 year) Compound Annual Growth Rate (n= 2045)** | **$212** | $266 | $ 3,259 | | |
| **Estimated Costs (per business over 10-year Period) Compound Annual Growth Rate (n=2045)** | **$0.10** | $0.13 | $1.59 | | |
| **Estimated Costs (annually) Compound Annual Growth Rate (n=2045)** | **$21** | $26 | $325 | | |

| | |
|---|---|
| UK OSA in scope business | 25000 |
| In scope Australian business | 2045 |
| | 8.18 |
| %Difference | 0.0818 |

#### 9.2.2.15. Estimating quantifiable harms

As set out above, any attempt to estimate the monetary costs of abuse is reductive to victim-survivors. This analysis is not intended to diminish the terrible impacts experienced by victim-survivors – any financial qualifications of harm can never represent the considerable and unmeasurable human costs of abuse.

The key assumptions in relation to this part of the required assessment are that the demographics and the population that would be impacted by the policy options considered here are transferrable between the study locations. The studies used as sources for the cost estimates are as follows:

- Letourneau EJ, Brown DS, Fang X, Hassan A, Mercy JA. The economic burden of child sexual abuse in the United States, *Journal Child Abuse Neglect*, 2018 May; 79, pp 413-422. (United States 2018)
- Saied-Tessier, A. (2014). Estimating the costs of child sexual abuse in the UK, *National Society for the Prevention of Cruelty to Children* (United Kingdom 2014)

There are considerable variations in the timeframe, scope, and methodologies used for each of the studies and they are not considered to be equally comparable. For example, the UK OSA measured 'contact' child sexual abuse and it is not clear if the United States (2018) and United Kingdom (2014) studies differentiated between these types of offending. These studies were selected because they explicitly costed the child sexual abuse, rather than available studies that costed more broader harms such as child sexual abuse, maltreatment, and neglect. For accuracy and comparability these broader studies were omitted from the analysis.

**Assumptions**

While extrapolation and application of findings from the cited studies cannot be directly applied to the Australian context without considering adjustment for differences in health care, welfare, job markets, offence reporting, criminal justice, population size and education systems (in the absence of any Australian studies estimating costs of online child sexual abuse) these are indirectly used as representative estimates that could reasonably expected to be similar costs in Australia.

A key limitation of the studies used is that they are dated. The United Kingdom study is based on 2011 prevalence data and the United States study uses 2015 data. To determine the costs in 2023, the data period for each study was adjusted for United States[64] and United Kingdom[65] inflation rates as of 2023. This was to ensure that the rates presented were contemporary, however this does not factor in differences in inflation in the United Kingdom and United States and how this varies from Australia. This is to provide an indication only.

While these costs are significant, it is reiterated that the burden of 'online' child sexual abuse is unlikely however to all be linked to RES and DIS and the exact proportion that could be attributed to RES and DIS cannot be estimated.

Due to the absence of available research material to draw from, the estimated quantified harms are limited to child sexual abuse only and does not capture the costs that would be incurred on individuals, community and government from the access, production, and distribution of other harmful material, such as pro-terror and extreme violence on RES and DIS.

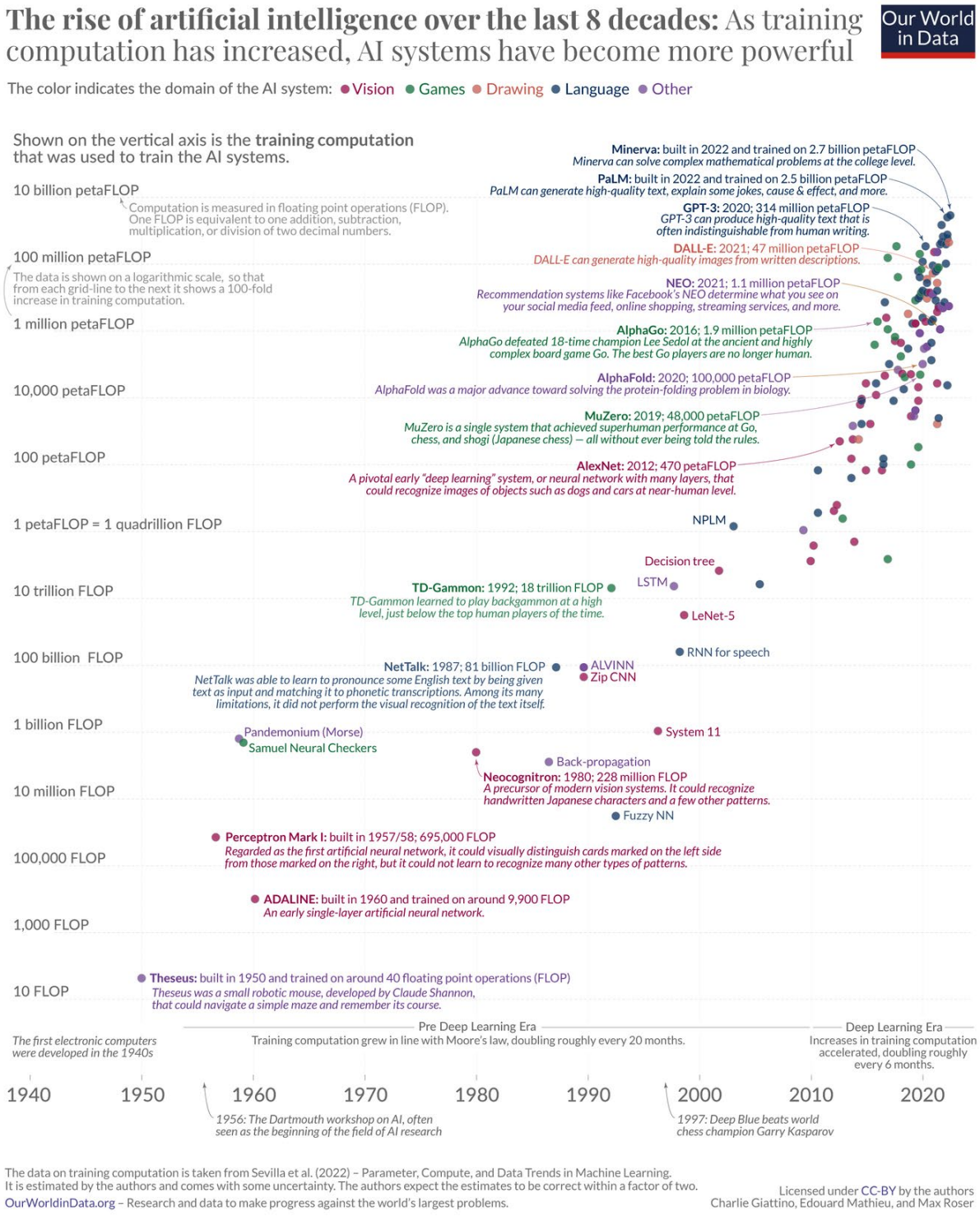
**Table 14 – Online Child Sexual Abuse and Child Sexual Abuse Studies – estimated annual costs**

| *Offence Nature* | Online Child Sexual Abuse | *Child Sexual Abuse* | *Child Sexual Abuse* |
|---|---|---|---|
| *Jurisdiction where costs estimated* | United Kingdom | United Kingdom | United States |
| *Date of Study/Data* | 2021-2022 | 2011 | 2015 |
| *Cost annual (source figures)* | 933 million pounds (0.99 billion) | 3 billion pounds | $9.3 billion USD |
| *Current Estimated Annual Cost (Adjusted Inflation 2023)* | 1.1 billion pounds | 4.3 billion pounds | $12 billion USD |
| *Estimated Annual Cost in AUD* | A$2.1 billion | A$8.2 billion | A$18.4 billion |

[64] Inflation Calculator | Find US Dollar's Value From 1913-2024 (usinflationcalculator.com)
[65] Inflation calculator | Bank of England

| *Offence Nature* | Online Child Sexual Abuse | *Child Sexual Abuse* | *Child Sexual Abuse* |
|---|---|---|---|
| *Source* | UK Department for Digital, Culture, Media and Sport. (2022, January 31). Online Safety Bill: Impact assessment. London, United Kingdom | Saied-Tessier, A. (2014). Estimating the costs of child sexual abuse in the UK. National Society for the Prevention of Cruelty to Children NSPCC library catalogue United Kingdom (2014) | Letourneau EJ, Brown DS, Fang X, Hassan A, Mercy JA. The economic burden of child sexual abuse in the United States. Child Abuse Neglect. 2018 May; 79 :413-422. The economic burden of child sexual abuse in the United States - PubMed (nih.gov) United States (2018) |

# 9.3. Annexure C - The rise of artificial intelligence over the last 8 decades: As training computation has increased, AI systems have become more powerful.



(Roser, 2022)

## 9.4. Annexure D – Risk categories for RES and DIS providers.

The Standards recognises the different functionalities, risk and capabilities of services and sets out specific requirements for particular categories.

If a RES or DIS does not fall within a category as defined in the standard and further outlined below, the service would need to undertake a risk assessment and would be classified in one of the following:

- Tier 1 RES/DIS: high risk
- Tier 2 RES/DIS: medium risk
- Tier 3 RES/DIS: low risk

**Defined and pre-assessed categories and risk tiers for RES Standard**

| Specific categories | Description |
|---|---|
| **Communication relevant electronic service** | This includes services that enable a user to communicate with another user and view, navigate or search for other users with, or without, already having their contact details which does not fit the other categories in the RES Standard (i.e. online messaging services and some video conferencing services, as well as some carriage services (email but not text messaging)). |
| **Gaming service with communication functionality** | A service that enables end-users to play online games with each other and share material with each other (for example, URLs, hyperlinks, images and/or videos). |
| **Gaming service with limited communication functionality** | A service that enables end-users to play online games with each other but only allows limited sharing of material (for example, in-game images and/or pre-selected messages). |
| **Dating service** | A service primarily used for dating that has a messaging function. This category does not include escort or sex work services |

The RES Standard also identifies a group of defined categories of relevant electronic services, which also have specific requirements under the RES Standard.

**Defined categories**

| Category | Description |
|---|---|
| Telephony RES | A Short Message Service (SMS) or Multimedia Messaging Service (MMS) provided over a public mobile telecommunications service |
| Enterprise RES | A service being provided to an organisation to enable people within that organisation to communicate with each other. |

### Categories in the DIS Standard

**Specific categories of DIS**

**End-user-managed hosting service**: an online service primarily designed or adapted to enable end-users to store or manage material e.g., cloud storage for files/photos.

**Generative AI categories:**

- **High impact generative AI DIS:** an online service that uses machine learning models to enable an end-user to generate synthetic high impact (X 18+ or RC) material. E.g., nudify apps and pornography generators.
- **Model distribution platform:** an online service which allows end-users to upload machine learning models, and which makes models available for download by other end-users.

**Tiered categories for other DIS**

| Tier | Description |
|---|---|
| Tier 1 | **High impact DIS**, a website or app (which is not a social media or relevant electronic service) that has the sole or predominant purpose of enabling access to high impact material[66] posted by users. E.g., 'gore' sites, pornography sites. |
| Tier 2 | A website or app which is not a social media or relevant electronic service, is not Tier 1, Tier 3 service or otherwise fall within a defined or pre-assessed category. By way of example, an online service which makes available professionally produced material and end-user generated material, and where posted material is only visible to known users. |

[66] High impact material, is defined in the DIS Standard as: films or computer games which have been or, if classified, would be classified R18+, X18+ or Refused Classification (RC) in accordance with *the Classification Act 1995*; and publications which have been or, if classified, would likely be classified Category 1 Restricted, Category 2 Restricted, or RC in accordance with the *Classification Act 1995*.

| Tier | Description |
|---|---|
| Tier 3 | **Classified DIS**, e.g., websites providing general entertainment that would be classified as R18+ or lower.<br>**General Purpose DIS**, websites or apps which provide general information e.g., news, educational and health websites.<br>**Enterprise DIS**, services provided to an organisation for use in the organisation's activities. |

# 10. References

Australian Bureau of Statistics. (2023, August 22). *Counts of Australian Businesses, including Entries and Exits*. Retrieved from ABS Web site: https://www.abs.gov.au/statistics/economy/business-indicators/counts-australian-businesses-including-entries-and-exits/latest-release

Australian Centre to Counter Child Exploitation. (2021). *ACCCE Statistics*. Retrieved from Australian Centre to Counter Child Exploitation Web site: https://www.accce.gov.au/resources/research-and-statistics/statistics

Australian Child Maltreatment Study. (2023). *The prevalence and impact of child maltreatment in Australia: Findings from the Australian Child Maltreatment Study: 2023 Brief Report*. Brisbane: Queensland University of Technology.

Australian Communications and Media Authority and eSafety Commissioner. (2023). *Annual Report 2022-23*. Canberra: Australian Communications and Media Authority.

Australian Federal Police. (2021, September 11). *AFP warn about fast growing online child abuse trend*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/afp-warn-about-fast-growing-online-child-abuse-trend#:~:text=Australian%20children%20as%20young%20as,producing%20even%20more%20graphic%20content.

Australian Federal Police. (2022, March 2). *Nationwide Operation Molto closes with the removal of 51 children from harm in Australia*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/nationwide-operation-molto-closes-removal-51-children-harm-australia

Australian Federal Police. (2023, January 28). *AFP urges parents and carers to be vigilant about online gaming*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/afp-urges-parents-and-carers-be-vigilant-about-online-gaming#:~:text=%E2%80%9CAs%20part%20of%20the%20grooming,sending%20them%20child%20abuse%20material.%E2%80%9D

Australian Federal Police. (2023, December 3). *Holiday season warning: Extremists infiltrating online and gaming platforms to recruit young Australians*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/holiday-season-warning-extremists-infiltrating-online-and-gaming

Australian Federal Police. (2023, July 17). *SA man jailed for transmitting child abuse material*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/sa-man-jailed-transmitting-child-abuse-material

Australian Government Attorney-General's Department. (2015). *Preventing violent extremism and radicalisation in Australia.* Canberra: Australian Government Attorney-General's Department.

Australian Institute of Criminology. (2022). *Crime & justice research 2022: Online sexual exploitation of children.* Canberra: Australian Institute of Criminology.

Australian Institute of Criminology. (2022). *Crime & Justice Research Report 2022: Online Sexual Exploitation of Children.* Canberra.

Australian Institute of Family Studies. (2018, September). *The economic costs of child abuse and neglect*. Retrieved from Australian Institute of Family

Studies Web site: https://aifs.gov.au/resources/policy-and-practice-papers/economic-costs-child-abuse-and-neglect

Australian Institute of Health and Welfare. (2024, February 29). *Illicit drug use*. Retrieved from Australian Institute of Health and Welfare Web site: https://www.aihw.gov.au/reports/illicit-use-of-drugs/illicit-drug-use

Australian Security Intelligence Organisation. (2019). *Counter-terrorism*. Retrieved from Australian Security Intelligence Organisation Web site: https://www.asio.gov.au/about/what-we-do/counter-terrorism

Autralian Bureau of Statistics. (2018, March 28). *Household use of information technology*. Retrieved from Australian Bureau of Statistics Web site: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/latest-release

Binder, J. F., & Kenyon, J. (2022). Terrorism and the internet: How dangerous is online radicalization? *Front Psychol*. doi:10.3389/fpsyg.2022.997390

Bravehearts. (2024). *Technology facilitated abuse*. Retrieved from Bravehearts Web site: https://bravehearts.org.au/research-lobbying/stats-facts/online-risks-child-exploitation-grooming/

Canadian Centre for Child Protection. (2021). *Project Arachnid: Online availability of child sexual abuse material*. Winnipeg: Canadian Centre for Child Protection.

Canadian Centre for Child Protection Inc. (2017). *SURVIVORS' SURVEY: FULL REPORT 2017*. Winnipeg: Canadian Centre for Child Protection Inc.

Canadian Centre for Child Protection Inc. (2023). *PARENTS' PERSPECTIVES ON HOW CHILD SEXUAL ABUSE MATERIAL IMPACTS THE ENTIRE FAMILY: System Failures, Resilience, and Recommendations for Change* . Winnipeg: Canadian Centre for Child Protection Inc.

Child Rescue Coalition. (2024). *AMELIA'S STORY: GROOMED ON A VIDEO GAME*. Retrieved from Child Rescue Coalition Web site: https://childrescuecoalition.org/educations/amelias-story-groomed-on-a-video-game/

Clark, N. (2023, April 19). 'Grandma exploit' tricks Discord's AI chatbot into breaking its own ethical rules. *Polygon*. Retrieved from https://www.polygon.com/23690187/discord-ai-chatbot-clyde-grandma-exploit-chatgpt

Commission for Countering Extremism. (2020). *COVID-19: How hateful extremists are exploiting the pandemic*. London: Commission for Countering Extremism.

Counter Extremism Project. (2018, August 10). *Terror Group's Use Of Cloud Storage Supplements Acts Of Violence & Brutality*. Retrieved from Counter Extremsim Project Web site: https://www.counterextremism.com/press/isis%E2%80%99-tactics-shift-battlefield-over-internet

Counter Extremism Project. (2024). *Terrorists on Telegram*. New York: Counter Extremism Project.

Crawford, A., & Smith, T. (2023, June 29). Illegal trade in AI child sex abuse images exposed. *BBC News*. Retrieved from https://www.bbc.com/news/uk-65932372

Crisan, L. (2023, December 6). *Launching Default End-to-End Encryption on Messenger*. Retrieved from Meta Web site: https://about.fb.com/news/2023/12/default-end-to-end-encryption-on-messenger/#:~:text=Takeaways,media%20quality%20and%20disappearing%20messages

CSIRO. (2024, March 27). *Artificial Intelligence Foundation Models Report*. Retrieved from CSIRO: https://www.csiro.au/en/research/technology-space/ai/ai-foundation-models-report

Davis, A., & Rosen, G. (2019, August 1). *Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer*. Retrieved from Meta web site: https://about.fb.com/news/2019/08/open-source-photo-video-matching/

Department of Home Affairs. (2020, March). Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. Canberra. Retrieved from https://www.homeaffairs.gov.au/news-subsite/files/voluntary-principles-counter-online-child-sexual-exploitation-abuse.pdf

Drejer, C., Riegler, M. A., Halvorsen, P., Baugerud, G. A., & Johnson, M. S. (2023). Livestreaming Technology and Online Child Sexual Exploitation and Abuse: A Scoping Review,. *Trauma, Violence, & Abuse, 25*(1). doi:10.1177/15248380221147564

ECPAT International; INTERPOL; UNICEF. (2024). *Disrupting Harm*. Retrieved from ECPAT Web site: https://ecpat.org/disrupting-harm/

eSafety Commissioner. (2018). *Digital families*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/research/digital-parenting/digital-families

eSafety Commissioner. (2021). *Development of industry codes under the Online Safety Act: Position Paper*. Canberra: eSafety Commissioner.

eSafety Commissioner. (2021, November 11). *Supervising preschoolers online*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/research/digital-parenting/supervising-preschoolers-online

eSafety Commissioner. (2022). *Basic Online Safety Expectations: Summary of industry responses to the first mandatory transparency notices.* Canberra: eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf

eSafety Commissioner. (2022). *Mind the Gap - Paremtal awareness of children's exposure to risks online*. Canberra: eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf

eSafety Commissioner. (2023, May 31). Summary of Reasons – Designated Internet Services Code. Canberra.

eSafety Commissioner. (2023, May 31). Summary of Reasons – Relevant Electronic Services Code. Canberra.

eSafety Commissioner. (2023). *Tech Trends Position Statement: Generative AI*. Canberra: eSafety Commissioner.

eSafety Commissioner. (2023, October 17). *Updated Position Statement: End-to-end encryption*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/sites/default/files/2023-10/End-to-end-encryption-position-statement-oct2023.pdf

eSafety Commissioner. (2024). *Levelling up to stay safe: Young people's experiences navigating the joys and risks of online gaming.* Canberra: eSafety Commissioner.

European Commission. (2024, March 24). *AI Act*. Retrieved from European Commission Web site: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20AI%20act%20aims%20to,%2Dsized%20enterprises%20(SMEs).

European Commission. (2024). *The Digital Services Act*. Retrieved from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

Europol. (2023). *ChatGPT: The impact of Large Language Models on Law Enforcement.* Luxembourg: European Union Agency for Law Enforcement Cooperation. Retrieved from https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf

Farid, H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety, 1*(4). Retrieved from https://doi.org/10.54501/jots.v1i4.56

Fenech, S. (2023, March 15). More than a third of kids under 12 already own a smartphone, new research reveals. *TechGuide*. Retrieved from https://www.techguide.com.au/news/mobiles-news/more-than-a-third-of-kids-under-12-already-own-a-smartphone-new-research-reveals/

Fitzsimmons, C. (2021, October 19). Australia among the worst for online sexual harm to children. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/national/australia-among-the-worst-for-online-sexual-harm-to-children-20211018-p590xt.html

Gewirtz-Meydan, A., Walsh, W., Wolak, J., & Finkelhor, D. (2018). The complex experience of child pornography survivors. *Child Abuse & Neglect*(80), 238-248. doi:10.1016/j.chiabu.2018.03.031

Giles, S., Alison, L., Humann, M., Tejeiro, R., & Rhodes, H. (2024). Estimating the economic burden attributable to online only child sexual abuse offenders: implications for police strategy. *Frontiers in Psychology, 14*. Retrieved from https://doi.org/10.3389/fpsyg.2023.1285132

Global Internet Forum to Counter Terrorism. (2024). *Technical Products*. Retrieved from GIFCT Web site: https://gifct.org/tech/

Global Online Safety Regulators Network. (2022). Terms of Reference 2022-23. eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-11/Terms%20of%20Reference%20-%20%20The%20Global%20Online%20Safety%20Regulators%20Network.pdf

Global Online Safety Regulators Network. (2023, September). Position Statement: Human RIghts and Online Safety Regulation. eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2023-09/Position-statement-Human-rights-and-online-safety-regulation.pdf

Google. (2024). *Fighting child sexual abuse online*. Retrieved from Google Web site: https://protectingchildren.google/#tools-to-fight-csam

Graham, A., & Sahlberg, P. (2021). *Growing Up Digital Australia: Phase 2 Technical Report.* Sydney: UNSW Gonski Institute for Education.

Hardy, J., & Stewart, C. (2023, June 29). *Gore and violent extremism: How extremist groups exploit 'gore' sites to view and share terrorist material*. Retrieved from Institute for Strategic Dialogue: https://www.isdglobal.org/digital_dispatches/gore-and-violent-extremism-how-extremist-groups-exploit-gore-sites-to-view-and-share-terrorist-material/

Harriet, G. (2021, September 27). Online child abuse survey finds third of viewers attempt contact with children. *The Guardian*. Retrieved from https://www.theguardian.com/global-development/2021/sep/27/online-child-abuse-survey-finds-third-of-viewers-attempt-contact-with-children

Hart, M., Davey, J., Maharasingam-Shah, E. O., & Gallagher, A. (2021). *An Online Environmental Scan of Right-wing Extremism in Canada.* London: Institute for Strategic Diaglogue.

Insoll, T., Ovaska, A. K., & Nurmi, J. A.-V. (2022). Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an Anonymous Multilingual Survey on the Dark Web. *Journal of Online Trust and Safety, 1*(2). doi:https://doi.org/10.54501/jots.v1i2.29

International Justice Mission. (2023, September 13). Children are Not for Sale: Examining the Threat of Exploitation of Children in the U.S. and Abroad. Arlington, Virginia, USA: International Justice Mission.

Internet Watch Foundation. (2022). *The Annual Report 2021.* Cambridge: Internet Watch Foundation.

ITU. (2024). *Facts and Figures 2023: Internet use*. Retrieved from ITU Web site: https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/

John, A. S. (2022, September 19). Trust and Safety services' market to reach $15-20 billion by 2024 as metaverse continues to evolve. *Wire19*, pp. https://wire19.com/trust-and-safety-services-market/.

Joleby, M., Lunde, C., Landstrom, S., & Jonsso, L. S. (2020). "All of Me Is Completely Different": Experiences and Consequences Among Victims of Technology-Assisted Child Sexual Abuse. *Frontiers in Psychology, 11*. doi:10.3389

Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., . . . Jernite, Y. (2024). On the Societal Impact of Open Foundation Models. *Stanford University*.

Kaur, N., Rutherford, C. G., Martins, S. S., & Keyes, K. M. (2020). Associations between digital technology and substance use among U.S. adolescents: Results from the 2018 Monitoring the Future survey. *Drug and Alcohol Dependence, 213*. Retrieved from https://doi.org/10.1016/j.drugalcdep.2020.108124

Llanos, T. (2022). *Transparency reporting on terrorist and violent extremist content online 2022.* Paris: OECD Publishing.

Morgan, S. (2024, February 1). The World Will Store 200 Zettabytes Of Data By 2025. *Cybercrime Magazine*. Retrieved from https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/

Napier, S., & Teunissen, C. (2023). Overlap between child sexual abuse live streaming, contact abuse and other forms of child exploitation. *Trends & issuesin crime and criminal justice, 671*. Retrieved from https://www.aic.gov.au/sites/default/files/2023-05/ti671_overlap_between_csa_live_streaming_contact_abuse_and_other_child_exploitation.pdf

Napier, S., Teunissen, C., & Boxall, H. (2021). Live streaming of child sexual abuse: An analysis of offender chat logs. *Trends & issuesin crime and criminal justice, 639*.

Napier, S., Teunissen, C., & Boxall, H. (n.d.). How do child sexual abuse live streaming offenders access victims? *Trends & issues in crime and criminal justice*(642). Retrieved from https://doi.org/10.52922/ti78474

National Center for Missing & Exploited Children. (2023). *2022 CyberTipline Reports by Electronic Service Providers (ESP).* Alexandria: National Center for Missing & Exploited Children.

National Crime Agency. (2023, November 8). *Assistant head teacher caught with 11,500 child abuse images*. Retrieved from National Crime Agency Web site: https://nationalcrimeagency.gov.uk/news/assistant-head-teacher-caught-with-11-500-child-abuse-images

NetClean. (2021). *The NetClean Report: COVID-19 IMPACT 2020.* Göteborg: NetClean. Retrieved from https://www.datocms-assets.com/74356/1662373830-netcleanreport-2020.pdf

New Zealand Customs Service. (2023, September 4). *Man faces jail following Customs investigations into online child exploitation*. Retrieved from New Zealand Customs Service Web site: https://www.customs.govt.nz/about-us/news/media-releases/man-faces-jail-following-customs-investigations-into-online-child-exploitation/?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news

Nilsson, M. G., Tzani-Pepelasis, C., Ioannou, M., & Lester, D. (2019). Understanding the link between Sextortion and Suicide. *International Journal of Cyber Criminology, 13*(1). doi:10.5281/zenodo.3402356

Noroozian, A., Koenders, J., van Veldhuizen, E., Gana, C. H., Alrwais, S., McCoy, M., & van Eeten, M. (2019). Platforms in Everything: Analyzing Ground-Truth Data on the Anatomy and Economics of Bullet-Proof Hosting. *28th USENIX Security Symposium.* Santa Clara: USENIX. Retrieved from https://www.usenix.org/conference/usenixsecurity19/presentation/noroozian

NSPCC. (2020). *The impact of the coronavirus pandemic on child welfare: online abuse.* London: NSPCC.

NSW Finance, Services and Innovation. (2016, October). *Guidance for regulators to implement outcomes and risk-based regulation.* Retrieved from NSW Productivity Commission Web site: https://www.productivity.nsw.gov.au/sites/default/files/2018-05/Guidance_for_regulators_to_implement_outcomes_and_risk-based_regulation-October_2016.pdf

NSW Police Force. (2024, February 22). *Ninth man charged over involvement in international child abuse ring - Strike Force Packer*. Retrieved from NSW Police Force Web site: https://www.police.nsw.gov.au/news/news_article?sq_content_src=%2BdXJsPWh0dHBzJTNBJTJGJTJGZWJpenByZC5wb2xpY2UubnN3Lmdvdi5hdSUyRm1lZGlhJTJGMTEwNTY1Lmh0bWwmYWxsPTE%3D

OECD. (2021). *Risk-based regulation: Making sure that rules are science-based, targeted, effective and efficient.* Retrieved from OECD Web site: https://www.oecd.org/gov/regulatory-policy/chapter-six-risk-based-regulation.pdf

OECD. (2022). Transparency reporting on terrorist and violent extremist content online 2022. *OECD Digital Economy Papers, 334*. Retrieved from https://doi.org/10.1787/a1621fc3-en.

OECD. (2023). Transparency reporting on child sexual exploitation and abuse online. *OECD Digital Economy Papers*(357). Retrieved from https://doi.org/10.1787/554ad91f-en

Ofcom. (2024). *Understanding Pathways to Online Violent Content Among Children.* London: Ofcom.

Office of the New York State Attorney General Letitia James. (2022). *Investigative Report on the role of online platforms in the tragic mass shooting in Buffallo on May 14, 2022.* New York: New York State Attorney General.

Online Safety Act 2023. (2023, October 26). Retrieved from https://bills.parliament.uk/bills/3137

Prescott, A., Sargent, J. D., & . Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time. *Psychological and cognitive sciences, 115*(40). Retrieved from https://www.pnas.org/doi/full/10.1073/pnas.1611617114

Rape, Abuse & Incest National Network. (2022, August 25). *What is Child Sexual Abuse Material (CSAM)*. Retrieved from RAINN Web site: https://www.rainn.org/news/what-child-sexual-abuse-material-csam

Rizoiu, M.-A., & Schneider, P. (2023, August 15). Research Confirms Human Moderators Can Curb Online Harm. *The Mirage*. Retrieved from https://www.miragenews.com/research-confirms-human-moderators-can-curb-1065322/

Roser, M. (2022, December 6). *The brief history of artificial intelligence: The world has changed fast – what might be next?* Retrieved from Out World in Data Web site: https://ourworldindata.org/brief-history-of-ai

Rowland MP, M. (2023, November 22). Address to the National Press Club. Canberra. Retrieved from The Hon Michelle Rowland MP: https://minister.infrastructure.gov.au/rowland/speech/address-national-press-club

Salter, M., & Sokolov, S. (2023). "Talk to strangers!" Omegle and the political economy of technology-facilitated child sexual exploitation. *Journal of Criminology, 57*(1). doi:https://doi.org/10.1177/2633807623119445

Schultz, A. (2023, December 3). Roblox used by extremists to recruit children, police warn. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/technology/video-games/roblox-used-by-extremists-to-recruit-children-police-warn-20231202-p5eohy.html

Suojellaan Lapsia, Protect Children ry. (2024). *Tech Platforms Used by Online Child Sexual Abuse Offenders: Research Report with Actionable Recommendations for the Tech Industry*. Helsinki: Suojellaan Lapsia, Protect Children ry.

Tech Against Terrorism. (2021). *GIFCT Technical Approaches Working Group: Gap Analysis and Recommendations for deploying technical solutions to tackle the terrorist use of the internet*. Global Internet Forum to Counter Terrorism.

Tech Against Terrorism. (2022). *State of Play 2022: Trends in Terrorist and Violent Extremist Use of the Internet.* London: Tech Against Terrorism.

Teunissen, C., & Napier, S. (2022). Child sexual abuse material and end-to-end encryption on social media platforms: An overview. *Trends & issues in crime and criminal justice*(653). Retrieved from https://www.aic.gov.au/sites/default/files/2022-07/ti653_csam_and_end-to-end_encryption_on_social_media_platforms.pdf

Teunissen, C., Thomsen, D., Napier, S., & Boxall, H. (2024). Risk factors for receiving requests to facilitate child sexual exploitation and abuse on dating apps and websites. *Trends & issuesin crime and criminal justice, 686*. Retrieved from https://doi.org/10.52922/ti77291

The Australian Centre to Counter Child Exploitation. (2020). *Online Child Sexual Exploitation: Understanding Community Awareness, Perceptions, Attitudes and Preventative Behaviours.* Brisbane: The Australian Centre to Counter Child Exploitation.

Thiel, D., Stroebel, M., & Portnoff, R. (2023). *Generative ML and CSAM: Implications and Mitigations.* Stanford: Stanford Internet Observatory. Retrieved from https://stacks.stanford.edu/file/druid:jv206yg3793/20230624-sio-cg-csam-report.pdf

Thorn. (2024). *Safer Essential: API-based CSAM detection built by Thorn*. Retrieved from AWS Marketplace Web site: https://aws.amazon.com/marketplace/pp/prodview-dfwekn4bx4ake

UK Department for Digital, Culture, Media and Sport. (2022, January 31). Online Safety Bill: Impact assessment. London, United Kingdom.

WeProtect Global Alliance. (2023). *Analysis of the sexual threats children face online*. Retrieved from WeProtect Global Alliance: https://www.weprotect.org/global-threat-assessment-23/analysis-sexual-threats-children-face-online/

WeProtect Global Alliance. (2023). *Global Threat Assessment 2023.* WeProtect Global Alliance.

WhatsApp. (n.d.). *How WhatsApp Helps Fight Child Exploitation*. Retrieved from WhatsApp Web site: https://faq.whatsapp.com/5704021823023684

Whiteford, H. (2022). The Productivity Commission inquiry into mental health. *Aust N Z J Psychiatry, 56*(4). doi:10.1177/00048674211031159

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology, 60*(1), 93-117. Retrieved from https://doi.org/10.1093/bjc/azz049

Winston, A. (2024, March 13). There Are Dark Corners of the Internet. Then There's 764. *Wired*. Retrieved from https://www.wired.com/story/764-com-child-predator-network/

Australian Bureau of Statistics. (2023, August 22). *Counts of Australian Businesses, including Entries and Exits*. Retrieved from ABS Web site: https://www.abs.gov.au/statistics/economy/business-indicators/counts-australian-businesses-including-entries-and-exits/latest-release

Australian Centre to Counter Child Exploitation. (2021). *ACCCE Statistics*. Retrieved from Australian Centre to Counter Child Exploitation Web site: https://www.accce.gov.au/resources/research-and-statistics/statistics

Australian Child Maltreatment Study. (2023). *The prevalence and impact of child maltreatment in Australia: Findings from the Australian Child Maltreatment Study: 2023 Brief Report.* Brisbane: Queensland University of Technology.

Australian Communications and Media Authority and eSafety Commissioner. (2023). *Annual Report 2022-23.* Canberra: Australian Communications and Media Authority.

Australian Federal Police. (2021, September 11). *AFP warn about fast growing online child abuse trend*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/afp-warn-about-fast-growing-online-child-abuse-trend#:~:text=Australian%20children%20as%20young%20as,producing%20even%20more%20graphic%20content.

Australian Federal Police. (2022, March 2). *Nationwide Operation Molto closes with the removal of 51 children from harm in Australia*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/nationwide-operation-molto-closes-removal-51-children-harm-australia

Australian Federal Police. (2023, January 28). *AFP urges parents and carers to be vigilant about online gaming*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/afp-urges-parents-and-carers-be-vigilant-about-online-gaming#:~:text=%E2%80%9CAs%20part%20of%20the%20grooming,sending%20them%20child%20abuse%20material.%E2%80%9D

Australian Federal Police. (2023, December 3). *Holiday season warning: Extremists infiltrating online and gaming platforms to recruit young Australians*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/holiday-season-warning-extremists-infiltrating-online-and-gaming

Australian Federal Police. (2023, July 17). *SA man jailed for transmitting child abuse material*. Retrieved from Australian Federal Police Web site: https://www.afp.gov.au/news-centre/media-release/sa-man-jailed-transmitting-child-abuse-material

Australian Government Attorney-General's Department. (2015). *Preventing violent extremism and radicalisation in Australia.* Canberra: Australian Government Attorney-General's Department.

Australian Institute of Criminology. (2022). *Crime & justice research 2022: Online sexual exploitation of children.* Canberra: Australian Institute of Criminology.
Australian Institute of Criminology. (2022). *Crime & Justice Research Report 2022: Online Sexual Exploitation of Children.* Canberra.
Australian Institute of Family Studies. (2018, September). *The economic costs of child abuse and neglect*. Retrieved from Australian Institute of Family Studies Web site: https://aifs.gov.au/resources/policy-and-practice-papers/economic-costs-child-abuse-and-neglect
Australian Institute of Health and Welfare. (2024, February 29). *Illicit drug use*. Retrieved from Australian Institute of Health and Welfare Web site: https://www.aihw.gov.au/reports/illicit-use-of-drugs/illicit-drug-use
Australian Security Intelligence Organisation. (2019). *Counter-terrorism*. Retrieved from Australian Security Intelligence Organisation Web site: https://www.asio.gov.au/about/what-we-do/counter-terrorism
Autralian Bureau of Statistics. (2018, March 28). *Household use of information technology*. Retrieved from Australian Bureau of Statistics Web site: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/latest-release
Binder, J. F., & Kenyon, J. (2022). Terrorism and the internet: How dangerous is online radicalization? *Front Psychol*. doi:10.3389/fpsyg.2022.997390
Bravehearts. (2024). *Technology facilitated abuse*. Retrieved from Bravehearts Web site: https://bravehearts.org.au/research-lobbying/stats-facts/online-risks-child-exploitation-grooming/
Canadian Centre for Child Protection. (2021). *Project Arachnid: Online availability of child sexual abuse material*. Winnipeg: Canadian Centre for Child Protection.
Canadian Centre for Child Protection Inc. (2017). *SURVIVORS' SURVEY: FULL REPORT 2017*. Winnipeg: Canadian Centre for Child Protection Inc.
Canadian Centre for Child Protection Inc. (2023). *PARENTS' PERSPECTIVES ON HOW CHILD SEXUAL ABUSE MATERIAL IMPACTS THE ENTIRE FAMILY: System Failures, Resilience, and Recommendations for Change* . Winnipeg: Canadian Centre for Child Protection Inc.
Child Rescue Coalition. (2024). *AMELIA'S STORY: GROOMED ON A VIDEO GAME*. Retrieved from Child Rescue Coalition Web site: https://childrescuecoalition.org/educations/amelias-story-groomed-on-a-video-game/
Clark, N. (2023, April 19). 'Grandma exploit' tricks Discord's AI chatbot into breaking its own ethical rules. *Polygon*. Retrieved from https://www.polygon.com/23690187/discord-ai-chatbot-clyde-grandma-exploit-chatgpt
Commission for Countering Extremism. (2020). *COVID-19: How hateful extremists are exploiting the pandemic.* London: Commission for Countering Extremism.
Counter Extremism Project. (2018, August 10). *Terror Group's Use Of Cloud Storage Supplements Acts Of Violence & Brutality*. Retrieved from Counter Extremsim Project Web site: https://www.counterextremism.com/press/isis%E2%80%99-tactics-shift-battlefield-over-internet
Counter Extremism Project. (2024). *Terrorists on Telegram.* New York: Counter Extremism Project.
Crawford, A., & Smith, T. (2023, June 29). Illegal trade in AI child sex abuse images exposed. *BBC News*. Retrieved from https://www.bbc.com/news/uk-65932372
Crisan, L. (2023, December 6). *Launching Default End-to-End Encryption on Messenger*. Retrieved from Meta Web site:

https://about.fb.com/news/2023/12/default-end-to-end-encryption-on-messenger/#:~:text=Takeaways,media%20quality%20and%20disappearing%20messages

CSIRO. (2024, March 27). *Artificial Intelligence Foundation Models Report*. Retrieved from CSIRO: https://www.csiro.au/en/research/technology-space/ai/ai-foundation-models-report

Davis, A., & Rosen, G. (2019, August 1). *Open-Sourcing Photo- and Video-Matching Technology to Make the Internet Safer*. Retrieved from Meta web site: https://about.fb.com/news/2019/08/open-source-photo-video-matching/

Department of Home Affairs. (2020, March). Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse. Canberra. Retrieved from https://www.homeaffairs.gov.au/news-subsite/files/voluntary-principles-counter-online-child-sexual-exploitation-abuse.pdf

Drejer, C., Riegler, M. A., Halvorsen, P., Baugerud, G. A., & Johnson, M. S. (2023). Livestreaming Technology and Online Child Sexual Exploitation and Abuse: A Scoping Review,. *Trauma, Violence, & Abuse, 25*(1). doi:10.1177/15248380221147564

ECPAT International; INTERPOL; UNICEF. (2024). *Disrupting Harm*. Retrieved from ECPAT Web site: https://ecpat.org/disrupting-harm/

eSafety Commissioner. (2018). *Digital families*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/research/digital-parenting/digital-families

eSafety Commissioner. (2021). *Development of industry codes under the Online Safety Act: Position Paper.* Canberra: eSafety Commissioner.

eSafety Commissioner. (2021, November 11). *Supervising preschoolers online*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/research/digital-parenting/supervising-preschoolers-online

eSafety Commissioner. (2022). *Basic Online Safety Expectations: Summary of industry responses to the first mandatory transparency notices.* Canberra: eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf

eSafety Commissioner. (2022). *Mind the Gap - Paremtal awareness of children's exposure to risks online.* Canberra: eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-02/Mind%20the%20Gap%20%20-%20Parental%20awareness%20of%20children%27s%20exposure%20to%20risks%20online%20-%20FINAL.pdf

eSafety Commissioner. (2023, May 31). Summary of Reasons – Designated Internet Services Code. Canberra.

eSafety Commissioner. (2023, May 31). Summary of Reasons – Relevant Electronic Services Code. Canberra.

eSafety Commissioner. (2023). *Tech Trends Position Statement: Generative AI.* Canberra: eSafety Commissioner.

eSafety Commissioner. (2023, October 17). *Updated Position Statement: End-to-end encryption*. Retrieved from eSafety Commissioner Web site: https://www.esafety.gov.au/sites/default/files/2023-10/End-to-end-encryption-position-statement-oct2023.pdf

eSafety Commissioner. (2024). *Levelling up to stay safe: Young people's experiences navigating the joys and risks of online gaming.* Canberra: eSafety Commissioner.

European Commission. (2024, March 24). *AI Act*. Retrieved from European Commission Web site: https://digital-

strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20AI%20act%20aims%20to,%2Dsized%20enterprises%20(SMEs).

European Commission. (2024). *The Digital Services Act*. Retrieved from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en

Europol. (2023). *ChatGPT: The impact of Large Language Models on Law Enforcement.* Luxembourg: European Union Agency for Law Enforcement Cooperation. Retrieved from https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf

Farid, H. (2022). Creating, Using, Misusing, and Detecting Deep Fakes. *Journal of Online Trust and Safety, 1*(4). Retrieved from https://doi.org/10.54501/jots.v1i4.56

Fenech, S. (2023, March 15). More than a third of kids under 12 already own a smartphone, new research reveals. *TechGuide*. Retrieved from https://www.techguide.com.au/news/mobiles-news/more-than-a-third-of-kids-under-12-already-own-a-smartphone-new-research-reveals/

Fitzsimmons, C. (2021, October 19). Australia among the worst for online sexual harm to children. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/national/australia-among-the-worst-for-online-sexual-harm-to-children-20211018-p590xt.html

Gewirtz-Meydan, A., Walsh, W., Wolak, J., & Finkelhor, D. (2018). The complex experience of child pornography survivors. *Child Abuse & Neglect*(80), 238-248. doi:10.1016/j.chiabu.2018.03.031

Giles, S., Alison, L., Humann, M., Tejeiro, R., & Rhodes, H. (2024). Estimating the economic burden attributable to online only child sexual abuse offenders: implications for police strategy. *Frontiers in Psychology, 14*. Retrieved from https://doi.org/10.3389/fpsyg.2023.1285132

Global Internet Forum to Counter Terrorism. (2024). *Technical Products*. Retrieved from GIFCT Web site: https://gifct.org/tech/

Global Online Safety Regulators Network. (2022). Terms of Reference 2022-23. eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2022-11/Terms%20of%20Reference%20-%20%20The%20Global%20Online%20Safety%20Regulators%20Network.pdf

Global Online Safety Regulators Network. (2023, September). Position Statement: Human RIghts and Online Safety Regulation. eSafety Commissioner. Retrieved from https://www.esafety.gov.au/sites/default/files/2023-09/Position-statement-Human-rights-and-online-safety-regulation.pdf

Google. (2024). *Fighting child sexual abuse online*. Retrieved from Google Web site: https://protectingchildren.google/#tools-to-fight-csam

Graham, A., & Sahlberg, P. (2021). *Growing Up Digital Australia: Phase 2 Technical Report.* Sydney: UNSW Gonski Institute for Education.

Hardy, J., & Stewart, C. (2023, June 29). *Gore and violent extremism: How extremist groups exploit 'gore' sites to view and share terrorist material*. Retrieved from Institute for Strategic Dialogue: https://www.isdglobal.org/digital_dispatches/gore-and-violent-extremism-how-extremist-groups-exploit-gore-sites-to-view-and-share-terrorist-material/

Harriet, G. (2021, September 27). Online child abuse survey finds third of viewers attempt contact with children. *The Guardian*. Retrieved from

https://www.theguardian.com/global-development/2021/sep/27/online-child-abuse-survey-finds-third-of-viewers-attempt-contact-with-children

Hart, M., Davey, J., Maharasingam-Shah, E. O., & Gallagher, A. (2021). *An Online Environmental Scan of Right-wing Extremism in Canada.* London: Institute for Strategic Diaglogue.

Insoll, T., Ovaska, A. K., & Nurmi, J. A.-V. (2022). Risk Factors for Child Sexual Abuse Material Users Contacting Children Online: Results of an Anonymous Multilingual Survey on the Dark Web. *Journal of Online Trust and Safety, 1*(2). doi:https://doi.org/10.54501/jots.v1i2.29

International Justice Mission. (2023, September 13). Children are Not for Sale: Examining the Threat of Exploitation of Children in the U.S. and Abroad. Arlington, Virginia, USA: International Justice Mission.

Internet Watch Foundation. (2022). *The Annual Report 2021.* Cambridge: Internet Watch Foundation.

ITU. (2024). *Facts and Figures 2023: Internet use*. Retrieved from ITU Web site: https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/

John, A. S. (2022, September 19). Trust and Safety services' market to reach $15-20 billion by 2024 as metaverse continues to evolve. *Wire19*, pp. https://wire19.com/trust-and-safety-services-market/.

Joleby, M., Lunde, C., Landstrom, S., & Jonsso, L. S. (2020). "All of Me Is Completely Different": Experiences and Consequences Among Victims of Technology-Assisted Child Sexual Abuse. *Frontiers in Psychology, 11*. doi:10.3389

Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., . . . Jernite, Y. (2024). On the Societal Impact of Open Foundation Models. *Stanford University*.

Kaur, N., Rutherford, C. G., Martins, S. S., & Keyes, K. M. (2020). Associations between digital technology and substance use among U.S. adolescents: Results from the 2018 Monitoring the Future survey. *Drug and Alcohol Dependence, 213*. Retrieved from https://doi.org/10.1016/j.drugalcdep.2020.108124

Llanos, T. (2022). *Transparency reporting on terrorist and violent extremist content online 2022.* Paris: OECD Publishing.

Morgan, S. (2024, February 1). The World Will Store 200 Zettabytes Of Data By 2025. *Cybercrime Magazine*. Retrieved from https://cybersecurityventures.com/the-world-will-store-200-zettabytes-of-data-by-2025/

Napier, S., & Teunissen, C. (2023). Overlap between child sexual abuse live streaming, contact abuse and other forms of child exploitation. *Trends & issuesin crime and criminal justice, 671*. Retrieved from https://www.aic.gov.au/sites/default/files/2023-05/ti671_overlap_between_csa_live_streaming_contact_abuse_and_other_child_exploitation.pdf

Napier, S., Teunissen, C., & Boxall, H. (2021). Live streaming of child sexual abuse: An analysis of offender chat logs. *Trends & issuesin crime and criminal justice, 639*.

Napier, S., Teunissen, C., & Boxall, H. (n.d.). How do child sexual abuse live streaming offenders access victims? *Trends & issues in crime and criminal justice*(642). Retrieved from https://doi.org/10.52922/ti78474

National Center for Missing & Exploited Children. (2023). *2022 CyberTipline Reports by Electronic Service Providers (ESP).* Alexandria: National Center for Missing & Exploited Children.

National Crime Agency. (2023, November 8). *Assistant head teacher caught with 11,500 child abuse images*. Retrieved from National Crime Agency Web site:

https://nationalcrimeagency.gov.uk/news/assistant-head-teacher-caught-with-11-500-child-abuse-images

NetClean. (2021). *The NetClean Report: COVID-19 IMPACT 2020.* Göteborg: NetClean. Retrieved from https://www.datocms-assets.com/74356/1662373830-netcleanreport-2020.pdf

New Zealand Customs Service. (2023, September 4). *Man faces jail following Customs investigations into online child exploitation*. Retrieved from New Zealand Customs Service Web site: https://www.customs.govt.nz/about-us/news/media-releases/man-faces-jail-following-customs-investigations-into-online-child-exploitation/?utm_source=miragenews&utm_medium=miragenews&utm_campaign=news

Nilsson, M. G., Tzani-Pepelasis, C., Ioannou, M., & Lester, D. (2019). Understanding the link between Sextortion and Suicide. *International Journal of Cyber Criminology, 13*(1). doi:10.5281/zenodo.3402356

Noroozian, A., Koenders, J., van Veldhuizen, E., Gana, C. H., Alrwais, S., McCoy, M., & van Eeten, M. (2019). Platforms in Everything: Analyzing Ground-Truth Data on the Anatomy and Economics of Bullet-Proof Hosting. *28th USENIX Security Symposium.* Santa Clara: USENIX. Retrieved from https://www.usenix.org/conference/usenixsecurity19/presentation/noroozian

NSPCC. (2020). *The impact of the coronavirus pandemic on child welfare: online abuse.* London: NSPCC.

NSW Finance, Services and Innovation. (2016, October). *Guidance for regulators to implement outcomes and risk-based regulation.* Retrieved from NSW Productivity Commission Web site: https://www.productivity.nsw.gov.au/sites/default/files/2018-05/Guidance_for_regulators_to_implement_outcomes_and_risk-based_regulation-October_2016.pdf

NSW Police Force. (2024, February 22). *Ninth man charged over involvement in international child abuse ring - Strike Force Packer*. Retrieved from NSW Police Force Web site: https://www.police.nsw.gov.au/news/news_article?sq_content_src=%2BdXJsPWh0dHBzJTNBJTJGJTJGZWJpenByZC5wb2xpY2UubnN3Lmdvdi5hdSUyRm1lZGlhJTJGMTEwNTY1Lmh0bWwmYWxsPTE%3D

OECD. (2021). *Risk-based regulation: Making sure that rules are science-based, targeted, effective and efficient.* Retrieved from OECD Web site: https://www.oecd.org/gov/regulatory-policy/chapter-six-risk-based-regulation.pdf

OECD. (2022). Transparency reporting on terrorist and violent extremist content online 2022. *OECD Digital Economy Papers, 334*. Retrieved from https://doi.org/10.1787/a1621fc3-en.

OECD. (2023). Transparency reporting on child sexual exploitation and abuse online. *OECD Digital Economy Papers*(357). Retrieved from https://doi.org/10.1787/554ad91f-en

Ofcom. (2024). *Understanding Pathways to Online Violent Content Among Children.* London: Ofcom.

Office of the New York State Attorney General Letitia James. (2022). *Investigative Report on the role of online platforms in the tragic mass shooting in Buffallo on May 14, 2022.* New York: New York State Attorney General.

Online Safety Act 2023. (2023, October 26). Retrieved from https://bills.parliament.uk/bills/3137

Prescott, A., Sargent, J. D., & . Hull, J. G. (2018). Metaanalysis of the relationship between violent video game play and physical aggression over time.

*Psychological and cognitive sciences, 115*(40). Retrieved from https://www.pnas.org/doi/full/10.1073/pnas.1611617114

Rape, Abuse & Incest National Network. (2022, August 25). *What is Child Sexual Abuse Material (CSAM)*. Retrieved from RAINN Web site: https://www.rainn.org/news/what-child-sexual-abuse-material-csam

Rizoiu, M.-A., & Schneider, P. (2023, August 15). Research Confirms Human Moderators Can Curb Online Harm. *The Mirage*. Retrieved from https://www.miragenews.com/research-confirms-human-moderators-can-curb-1065322/

Roser, M. (2022, December 6). *The brief history of artificial intelligence: The world has changed fast – what might be next?* Retrieved from Out World in Data Web site: https://ourworldindata.org/brief-history-of-ai

Rowland MP, M. (2023, November 22). Address to the National Press Club. Canberra. Retrieved from The Hon Michelle Rowland MP: https://minister.infrastructure.gov.au/rowland/speech/address-national-press-club

Salter, M., & Sokolov, S. (2023). "Talk to strangers!" Omegle and the political economy of technology-facilitated child sexual exploitation. *Journal of Criminology, 57*(1). doi:https://doi.org/10.1177/26338076231194445

Schultz, A. (2023, December 3). Roblox used by extremists to recruit children, police warn. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/technology/video-games/roblox-used-by-extremists-to-recruit-children-police-warn-20231202-p5eohy.html

Suojellaan Lapsia, Protect Children ry. (2024). *Tech Platforms Used by Online Child Sexual Abuse Offenders: Research Report with Actionable Recommendations for the Tech Industry*. Helsinki: Suojellaan Lapsia, Protect Children ry.

Tech Against Terrorism. (2021). *GIFCT Technical Approaches Working Group: Gap Analysis and Recommendations for deploying technical solutions to tackle the terrorist use of the internet*. Global Internet Forum to Counter Terrorism.

Tech Against Terrorism. (2022). *State of Play 2022: Trends in Terrorist and Violent Extremist Use of the Internet*. London: Tech Against Terrorism.

Teunissen, C., & Napier, S. (2022). Child sexual abuse material and end-to-end encryption on social media platforms: An overview. *Trends & issues in crime and criminal justice*(653). Retrieved from https://www.aic.gov.au/sites/default/files/2022-07/ti653_csam_and_end-to-end_encryption_on_social_media_platforms.pdf

Teunissen, C., Thomsen, D., Napier, S., & Boxall, H. (2024). Risk factors for receiving requests to facilitate child sexual exploitation and abuse on dating apps and websites. *Trends & issuesin crime and criminal justice, 686*. Retrieved from https://doi.org/10.52922/ti77291

The Australian Centre to Counter Child Exploitation. (2020). *Online Child Sexual Exploitation: Understanding Community Awareness, Perceptions, Attitudes and Preventative Behaviours.* Brisbane: The Australian Centre to Counter Child Exploitation.

Thiel, D., Stroebel, M., & Portnoff, R. (2023). *Generative ML and CSAM: Implications and Mitigations.* Stanford: Stanford Internet Observatory. Retrieved from https://stacks.stanford.edu/file/druid:jv206yg3793/20230624-sio-cg-csam-report.pdf

Thorn. (2024). *Safer Essential: API-based CSAM detection built by Thorn*. Retrieved from AWS Marketplace Web site: https://aws.amazon.com/marketplace/pp/prodview-dfwekn4bx4ake

UK Department for Digital, Culture, Media and Sport. (2022, January 31). Online Safety Bill: Impact assessment. London, United Kingdom.

WeProtect Global Alliance. (2023). *Analysis of the sexual threats children face online*. Retrieved from WeProtect Global Alliance: https://www.weprotect.org/global-threat-assessment-23/analysis-sexual-threats-children-face-online/

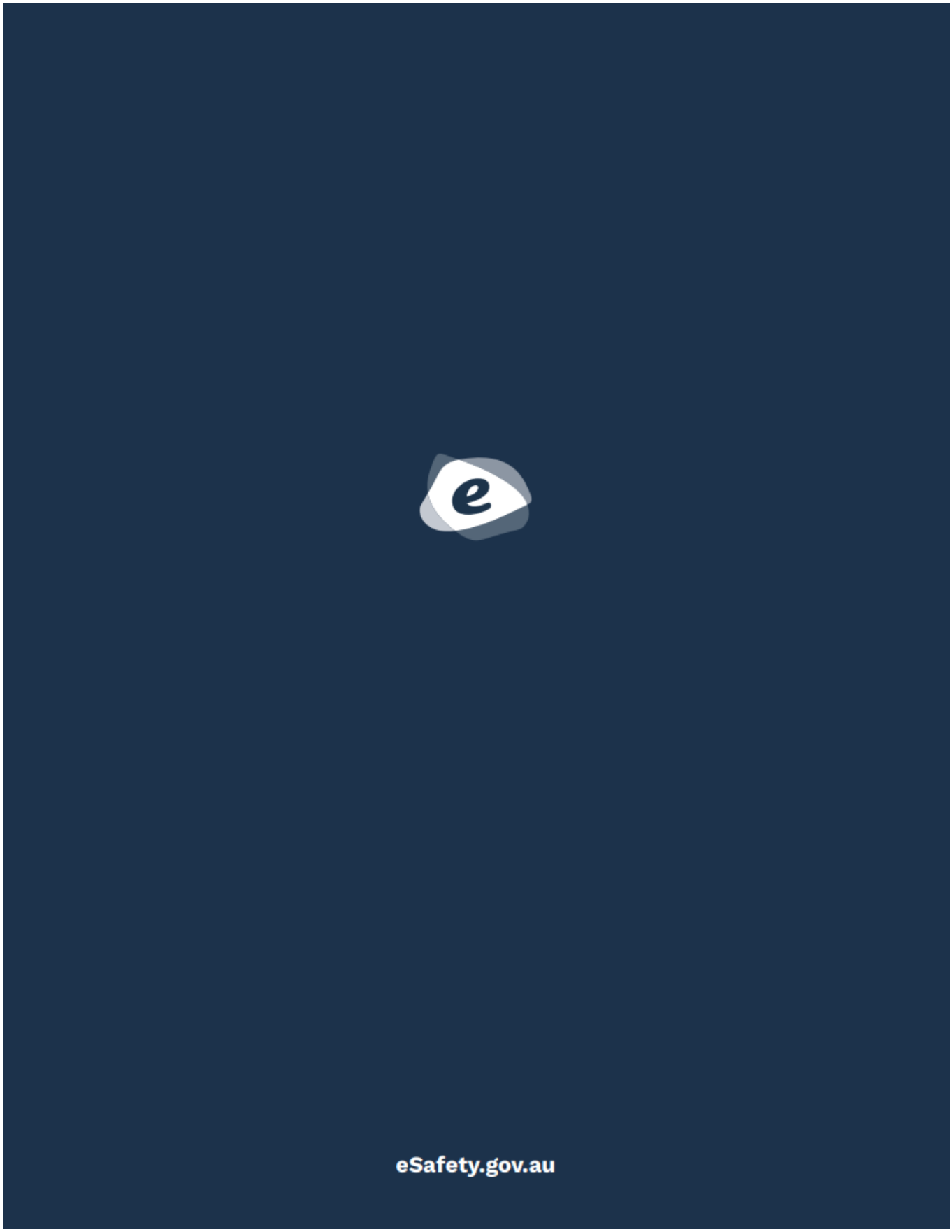
WeProtect Global Alliance. (2023). *Global Threat Assessment 2023.* WeProtect Global Alliance.

WhatsApp. (n.d.). *How WhatsApp Helps Fight Child Exploitation*. Retrieved from WhatsApp Web site: https://faq.whatsapp.com/5704021823023684

Whiteford, H. (2022). The Productivity Commission inquiry into mental health. *Aust N Z J Psychiatry, 56*(4). doi:10.1177/00048674211031159

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology, 60*(1), 93-117. Retrieved from https://doi.org/10.1093/bjc/azz049

Winston, A. (2024, March 13). There Are Dark Corners of the Internet. Then There's 764. *Wired*. Retrieved from https://www.wired.com/story/764-com-child-predator-network/


eSafety.gov.au