Australian Government
Department of Infrastructure, Transport,
Regional Development, Communications and the Arts

Statutory Review of the Online Safety Act 2021

Issues paper

April 2024

Content Warning: This paper discusses themes related to the prevention of online harms that may cause distress in readers. Themes include online and technology-facilitated abuse, sexual assault, child sexual exploitation and abuse, self-harm, suicide, eating disorders, pornographic content, and hateful language directed at groups of people. No case studies have been included.

Table of Contents

Key terms

Term	Meaning
Abhorrent violent conduct	Defined in section 474.32 of the Criminal Code. Includes a person engaging in a terrorist act, murder or attempted murder, or torture, rape or kidnapping of another person.
Abhorrent violent material	Defined in section 474.31 of the Criminal Code. Includes audio, visual or audio-visual material that records or streams abhorrent violent conduct engaged in by one or more persons, and that a reasonable person would regard in all the circumstances as being offensive. The material must also have been produced by a person or persons, each of whom is: • engaged in the abhorrent violent conduct, • conspired to engage in the abhorrent violent conduct, • aided, abetted, counselled, procured, or were in any way knowingly concerned in the abhorrent violent conduct, or • who attempted to engage in the abhorrent violent conduct. It is immaterial whether the material has been altered, or whether the abhorrent violent conduct was engaged in within or outside Australia.
App distribution service	Defined in section 5 of the *Online Safety Act 2021*. A service that enables end-users to download apps, where that download is by means of a carriage service. Examples include Apple App Store, and Google Play Store.
Basic Online Safety Expectations	The Basic Online Safety Expectations are determined under the *Online Safety Act 2021* and set out the Australian Government's expectations of the steps that should be taken by providers of social media services, messaging services, gaming services, apps and certain other sites accessible from Australia to keep Australians safe online. The *Online Safety Act 2021*provides eSafety with powers to require services to report on their compliance with the Basic Online Safety Expectations.
Caching service	A type of intermediary service that includes automatic, intermediate and temporary storage of information provided by a service recipient in a communication network for the purpose of making the onward transmission to other recipients on request more efficient. For example, a content delivery network (temporary storage or caching of files in geographically distributed servers to reduce the page loading time). (See Article 3 of the European Union's Digital Services Act).
Class 1 material – section 106 of the Online Safety Act 2021	Material that is or would likely be refused classification under Australia's National Classification Scheme, by reference to the National Classification Code. It includes material that: • depicts, expresses or otherwise deals with matters of sex, drug misuse or addiction, crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that they should not be classified

Term	Meaning
	• describes or depicts in a way that is likely to cause offence to a reasonable adult, a person who is, or appears to be, a child under 18 (whether the person is engaged in sexual activity or not), or • promotes, incites or instructs in matters of crime or violence. Class 1 material includes, for example, child sexual exploitation and abuse material and pro-terror material.
Class 2 material – section 107 of the Online Safety Act 2021	Material that is, or would likely be, classified under Australia's National Classification Scheme, by reference to the National Classification Code as either: • X18+ (or, in the case of publications, category 2 restricted), or • R18+ (or, in the case of publications, category 1 restricted), which is legally restricted to adults. Class 2 materials include, for example, pornography and other high impact material such as R18+ video games.
Designated internet service	Defined in section 14 of the *Online Safety Act 2021*, and only to the extent that material on the service is accessible to or delivered to one or more end-users in Australia. A service (other than a social media service, relevant electronic service, or on-demand program service) that allows end-users to access material on the internet using an internet carriage service or a service that delivers material to persons by means of an internet carriage service. This includes most apps and websites accessed by Australian end-users including retail websites, information apps (such as train timetables), and adult websites.
Hosting service	Defined in **Australia** as a service that hosts stored material that has been provided on a social media service, relevant electronic service, or designated internet service (see section 17 of the *Online Safety Act 2021*). Includes, for example Amazon Web Services. Defined in **Europe** as a type of intermediary service that stores information provided by a service recipient at their request. For example, cloud services. (See Article 3 of the European Union's Digital Services Act).
Internet carriage service	Defined in section 5 of the *Online Safety Act 2021* as 'a listed carriage service that enables end-users to access the internet.' The internet carriage service is provided to the public by an internet service provider. Examples include Optus, Telstra, and TPG Telecom Limited.
Internet search engine service	A service designed to collect, organise and/or rank material on the internet, that have the sole or primary purpose of allowing end-users to search the service's index of material for results in response to the end-user's queries, and the service returns search results in response to the query. Examples include Google Search, Microsoft Bing and Yahoo! Search.
Manufacturers, suppliers and installers of equipment	Definition applies where equipment is for use by end-users in Australia in connection with a social media service, relevant electronic service, designated internet service or internet carriage service. Examples include Apple and Samsung.

Term	Meaning
Material that depicts abhorrent violent conduct	Defined in section 9 of the *Online Safety Act 2021* as audio, visual or audio-visual material that records or streams abhorrent violent conduct. It is immaterial whether the material has been altered or who produced it.
Mere conduit service	A type of intermediary service that transmits information provided by a service recipient in a communication network, or provides access to a communication network. For example, virtual private networks, internet exchange points, or domain name system services. (See Article 3 of the European Union's Digital Services Act).
Relevant electronic service	Defined in section 13A of the *Online Safety Act 2021*, and only to the extent that material on the service is accessible or delivered to one or more end-users in Australia. A service that allows end-users to communicate with other end-users by means of email, instant messaging, short message service (SMS), multimedia message service (MMS), chat service or online game. Examples include Roblox, Gmail, and WhatsApp.
Sections of the online industry (defined in Part 1 of the Online Safety Act 2021)	Groups consisting of the providers of: • social media services • relevant electronic services • designated internet services • internet search engine services • app distribution services • hosting services • internet carriage services so far as those services are provided to end-users in Australia; and • the group consisting of persons who manufacture, supply, maintain or install equipment for use by end-users in Australia in connection with a social media service, relevant electronic service, designated internet service or internet carriage service.
Social media service	Defined in section 13 of the *Online Safety Act 2021*, and only to the extent that material on the service is accessible or delivered to one or more end-users in Australia. A service that has the sole or primary purpose of enabling online social interaction between end-users, where end-users can also link to other end-users and post material on the service. Examples include Facebook, Instagram, Tik Tok, and YouTube.

Part 1 – Introduction

Foreword

Australians have had access to the internet for over thirty years. In the 1990s, telecommunications was Australia's fastest growing industry and Australians were the second fastest in the world to take up this technology, behind the United States.

We are a nation of early adopters, using technology to drive commerce, education and research, access to health, social interaction, and information sharing. We know that communications and connectivity are drivers of productivity and can be a game-changer for families and small business, particularly in regional areas.

Much has changed since the 1990s, and digital technologies now provide us with opportunities and benefits that would have been unforeseeable at the inception of the internet. Online environments enable people to connect to communities, and proved indispensable in combatting the isolation presented by the COVID-19 pandemic.

But the proliferation of digital services and smart devices has posed challenges and facilitated harms that governments and communities around the world are now reckoning with - harms that often disproportionately impact the most vulnerable members of our society.

Indeed, the digital world is no longer unregulated or ungoverned. Australia is at the forefront of efforts to assert our values and expectations in the online environment. The office of the eSafety Commissioner is world-leading and plays a vital role in improving online safety for Australians. However, with rapid changes in technology and evolving community expectations, our regulatory and legislative frameworks must not be static.

That's why the Albanese Government brought forward the statutory review (the Review) of the *Online Safety Act 2021,* twelve months earlier than required – as digital platforms clearly need to do more to make their services safer than they are today.

Through the Review, being ably led by Ms Delia Rickard PSM, a former Deputy Chair of the Australian Competition and Consumer Commission for over a decade, the Government wants to hear from all parts of Australian society - parents, educators, industry, academics, and individuals of all ages - so we as a Government can clearly understand what the Australian public expects our online safety framework to accomplish.

The Review will consider optimal regulatory settings, and how to best support eSafety in reducing Australians' exposure to harms.

The need for flexibility so the eSafety Commissioner can respond to new and emerging technologies and harms is also vital. A prime example is generative AI that is now widely used at a scale far beyond what was possible when the Online Safety Act was passing through the Parliament in 2021.

Similarly, the role of social media being used to speak and amplify hate speech is of great concern to the Government, as is seriously harmful and illegal content such as child sexual abuse material and pro-terror content.

Governments around the world are collectively dealing with the question of effective digital platform regulation and how to address emerging risks and harms. International cooperation remains an important feature of internet governance, given the internet doesn't recognise international boundaries. We can learn a great deal from international approaches, many of which are canvassed in this paper, but we also have a lot to share.

Australia is a world-leader in online safety and will continue to play a leadership role in improving our laws and industry standards to contribute to growing international best practice.

Unfortunately, not every problem on the internet can be solved and not every harm can be eliminated. However, the task before us is to ensure community safeguards continue to improve, and digital platforms step up efforts to embed safety by design. Doing nothing - or going backwards - is not an option.

As the Minister for Communications, I am pleased to be able to commission this important and independent Review. I look forward to learning of its recommendations and encourage any member of the Australian public to take the opportunity to participate in this process.

The Hon Michelle Rowland MP
Minister for Communications

About the Review

On 22 November 2023, the Minister for Communications, the Hon Michelle Rowland MP, announced the commencement of a statutory review (the Review) into the operation of the *Online Safety Act 2021* (the Act).[1] Ms Delia Rickard PSM has been appointed to conduct the Review and provide a report of the Review to the Minister by 31 October 2024.

In the Government's April 2023 response to the House of Representatives Select Committee on Social Media and Online Safety Report, the Government committed to completing a statutory review of the Online Safety Act earlier than required under the Act, and within this term of Government, so that the Act can keep pace with the evolving online environment. This included a commitment to consider the operation of the existing framework of the Act and 'whether reforms are required to simplify regulatory arrangements including through the introduction of a duty of care requirement.'[2]

Australia has a strong track record in online safety. The Act commenced in January 2022 and introduced a world leading regulatory framework which at its core aims to improve and promote the safety of Australians online.

The Act introduced an adult cyber-abuse complaints scheme, formalised arrangements for dealing with material depicting abhorrent violent conduct, expanded the existing child cyberbullying scheme and strengthened the image-based abuse scheme. It introduced the first of its kind industry-based mechanisms, including greater transparency from online service providers around efforts to support user safety (Basic Online Safety Expectations), and provided for industry codes or standards to establish baseline requirements for the digital industry to address illegal and seriously harmful online content (Online Content Scheme).

The Act also empowered Australia's eSafety Commissioner to promote and improve online safety for Australians through a range of education, research, coordination, and advisory functions. The office of the eSafety Commissioner (eSafety) works with organisations and regulators domestically and around the world,[3] including as a founding member of the Global Online Safety Regulators Network (the only global forum currently dedicated to supporting collaboration between online safety regulators).

While digital technologies bring economic, educational and social opportunities, they can also present difficult and potentially harmful experiences. Online interactions (such as social media posts, direct messages, stories, or snaps) have expanded the vectors for harm and their ability to scale.[4] Recent advancements in technologies such as generative artificial intelligence are changing our online experiences, and can also be used to generate or amplify illegal and harmful content, including through the creation of synthetic child sexual exploitation and abuse material and other forms of harmful and extreme content.[5]

Online interactions have a broad impact on Australian lives. Australians work, study and access financial and professional services online, using a range of devices and services to engage with friends and family, interact with businesses and government, and to meet new people. Digital exclusion is increasingly a driver of

---

[1] The Hon Michele Rowland MP, Minister for Communications, 2023, Online Speeches, Address to the National Press Club 22 November Address to the National Press Club | Ministers for the Department of Infrastructure, accessed 12 February 2024.

[2] Australian Government (2023), Government Response to Social Media and Online Safety Report, 18, 20230330 - Australian Government response to the Social Media and Online Safety inquiry (infrastructure.gov.au)

[3] For more information refer to International engagement | eSafety Commissioner.

[4] The Hon Michele Rowland MP, Minister for Communications, 2023, Online Speeches, Address to the National Press Club 22 November Address to the National Press Club | Ministers for the Department of Infrastructure, accessed 12 February 2024.

[5] eSafety Commissioner, 2023, Tech Trends Position Statement – Generative AI, Generative AI - Position Statement - August 2023 .pdf (esafety.gov.au), accessed 12 February 2024.

inequality.[6] Choosing not to be online is no longer a practical option for most Australians. For many, the online world is a source of positivity and social connection.[7] However, in too many circumstances, Australians experience online harms resulting in a significant and detrimental effect on their lives.

In 2022, the House of Representatives Select Committee on Social Media and Online Safety reported hearing 'extensive evidence suggesting online harm is rampant on digital spaces,' with victims experiencing significant and lasting impacts ranging from psychological harm to impacts on career choices, or fears for personal safety.[8] The report identified several groups of Australians who were more likely to experience online harm or the effects of dangerous online behaviour. These included children, women (and women in public or prominent positions)[9], people from culturally or linguistically diverse backgrounds, Aboriginal and Torres Strait Islander peoples, people with particular religious beliefs, people who identify as LGBTQIA+, and older Australians.[10]

Australia is one of many countries regulating online safety in a global regulatory environment that is not confined to national borders. Since the Act commenced in 2022, other countries including Ireland, the European Union (EU), the United Kingdom (UK), Singapore and Sri Lanka have introduced online safety regulatory frameworks, evolving and learning from the rapidly changing regulatory environment. In the UK and EU (and as proposed in Canada) these frameworks have a focus on systems-based regulation, with enforceable requirements for digital service providers to conduct risk assessments and introduce mitigation measures.

Emerging global regulatory trends in online safety include systems-based regulation, transparency and accountability measures, risk assessments and mitigations, powers to compel corrective action, cost recovery provisions, significant penalties for non-compliance, as well as a focus on specific vectors for harm such as artificial intelligence. Acknowledging these recent and ongoing changes in the international regulatory environment, this Review offers an opportunity to consider whether Australia's regulatory framework would benefit from further alignment with other jurisdictions.

While the Government continues work to fully implement the Act, the global online safety landscape is rapidly evolving. The results and effectiveness of newer regulatory approaches are not yet fully understood. However, Australia needs to be responsive to global changes in regulating for a safer digital ecosystem and in how Australians connect and use digital products and services.

Review Terms of Reference

This Review will be a broad-ranging examination of the operation and effectiveness of the Act, including its existing regulatory schemes, penalties and enforcement, and any gaps in the Act. The Review will consider international developments in online safety regulation, including whether the law should be amended to impose a new duty of care on platforms towards their users.

---

[6] Australian Institute of Family Studies (2021), The digital divide in telepractice service delivery, September 2021, The digital divide in telepractice service delivery | Australian Institute of Family Studies (aifs.gov.au), accessed 26 April 2024.

[7] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), Social Media and Online Safety – Parliament of Australia (aph.gov.au), accessed 26 April 2024.

[8] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), 11. Social Media and Online Safety – Parliament of Australia (aph.gov.au), accessed 26 April 2024.

[9] Two-thirds of reports to eSafety about cyberbullying, image-based abuse and adult cyber abuse are made by women and girls. eSafety, How the new Online Safety Act supports women, OSA Fact sheet Women_0.pdf (esafety.gov.au), accessed 26 April 2024.

[10] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), 29-44, Social Media and Online Safety – Parliament of Australia (aph.gov.au), accessed 26 April 2024.

In addition, the Review's Terms of Reference list the following ten specific matters for consideration:

1. The overarching objects in section 3 of the Act, including the extent to which the objects and provisions of the Act remain appropriate to achieve the Government's current online safety policy intent.
2. The operation and effectiveness of the following statutory schemes and whether the regulatory arrangements should be amended:
    - cyber-bullying material targeted at an Australian child
    - non-consensual sharing of intimate images
    - cyber-abuse material targeted at an Australian adult
    - the Online Content Scheme, including the restricted access system and the legislative framework governing industry codes and standards, and
    - material that depicts abhorrent violent conduct.
3. The operation and effectiveness of the Basic Online Safety Expectations regime in the Act.
4. Whether additional arrangements are warranted to address online harms not explicitly captured under the existing statutory schemes, including:
    a. online hate
    b. volumetric (pile-on) attacks
    c. technology-facilitated abuse and technology-facilitated gender-based violence
    d. online abuse of public figures and those requiring an online presence as part of their employment
    e. other potential online safety harms raised by a range of emerging technologies, including but not limited to:
        - generative artificial intelligence
        - immersive technologies
        - recommender systems
        - end-to-end encryption
        - changes to technology models such as decentralised platforms
5. Whether the regulatory arrangements, tools and powers available to the Commissioner should be amended and/or simplified, including through consideration of:
    a. the introduction of a duty of care requirement towards users (similar to the United Kingdom's *Online Safety Act 2023* or the primary duty of care under Australia's work health and safety legislation) and how this may interact with existing elements of the Act
    b. ensuring industry acts in the best interests of the child
6. Whether penalties should apply to a broader range of circumstances.
7. Whether the current information gathering powers, investigative powers, enforcement powers, civil penalties or disclosure of information provisions should be amended.
8. The Commissioner's functions and governance arrangements, including:
    a. the Commissioner's roles and responsibilities under the Act
    b. whether the current functions and powers in the Act are sufficient to allow the Commissioner to carry out their mandate.
9. Whether the current governance structure and support arrangements for the Commissioner provided by the Australian Communications and Media Authority (ACMA) are fit for purpose for both the Commissioner and the ACMA.
10. Whether it would be appropriate to cost recover from industry for eSafety's regulatory activities.

Other Government actions

The *Online Safety Act 2021* is one part of the Government's broader suite of protections against online harms. Further detail on these other schemes can be found in Appendix 1.

Part 2 – Australia's regulatory approach to online services, systems and processes

> **Australia, alongside the rest of the world, must address the causes and amplifiers of harm to ensure a safe and equitable internet, especially to the most vulnerable members of society.**
>
> House of Representatives Select Committee on Social Media and Online Safety, *Social Media and Online Safety*, March 2022, [5.2].

Overview

The objects of the Act are to improve and promote the safety of Australians online.[11] Australia's regulatory approach includes schemes to provide corrective action to individuals in the case of specific types of harmful content alongside schemes aimed at improving service providers' online safety protections at a systemic level.

Regulated harms include child cyberbullying, adult cyber-abuse, the non-consensual sharing of intimate images, illegal and restricted content (Online Content Scheme), and material depicting abhorrent violent conduct. The Act makes online service providers more accountable for the online safety of Australians who use their services through the development and registration of enforceable industry codes and standards (under the Online Content Scheme) and through the Basic Online Safety Expectations regime.

Burden of responsibility for online safety

Submissions to the House of Representatives Select Committee on Social Media and Online Safety raised concerns that too much burden was placed on users taking responsibility for their personal safety online. The Committee reported that 'the time has come to fundamentally shift the burden of responsibility regarding ensuring online safety. For too long, the onus of maintaining online safety has been on the most vulnerable users, including children and their parents. This is unacceptable and unsustainable in an environment where users like children are exposed to the most risk online and suffer extreme forms of harm as a result.'[12]

The Online Safety Act's systems focus

The Act has two schemes that take a systems-focussed approach to preventing online harm: industry codes and standards created under the Online Content Scheme, and Basic Online Safety Expectations.

The Online Content Scheme provides for a broad spectrum of online services to be subject to industry codes or standards, with financial penalties for non-compliance with an industry standard and non-compliance with a direction to comply with an industry code. The focus of the codes and standards is limited to illegal or restricted material (by reference to the National Classification Code).

The Basic Online Safety Expectations establish minimum safety expectations for online service providers, but cover a narrower spectrum of services. The Commissioner can require regulated service providers to report on compliance with the expectations ('transparency reports'), with financial penalties applying for failure to comply with the required reporting. The expectations themselves do not create a legally enforceable duty.

---

[11] *Online Safety Act 2021*, section 3.

[12] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.78], Social Media and Online Safety – Parliament of Australia (aph.gov.au), accessed 26 April 2024.

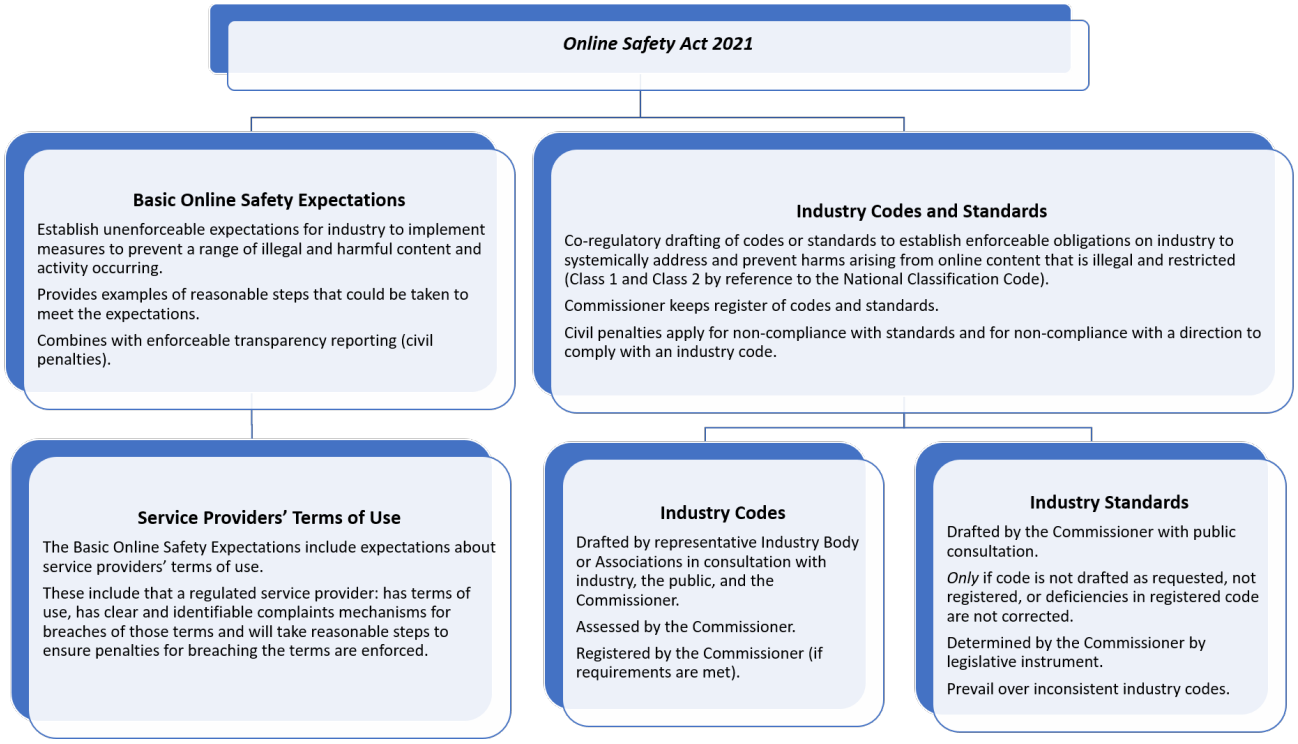

Figure 2.1 Overview of systems-based regulatory schemes in the Act.

Online Content Scheme – industry codes and standards

Under the Online Content Scheme, the Commissioner can register codes developed by industry bodies or associations representing sections of the online industry. Once registered, industry codes (and standards) are mandatory and enforceable, and take an outcomes-based approach to regulating Class 1 and Class 2 material.

The Act provides for industry bodies or associations representing a regulated section of industry to prepare draft codes to regulate 'Class 1' and 'Class 2' illegal and restricted online material. The Commissioner can register a code if it meets the statutory requirement of providing appropriate community safeguards. If not, the Commissioner may determine an industry standard. If both an industry code and an industry standard apply to the same person, the industry standard will prevail to the extent of any inconsistency.[13] If there are no representative industry bodies or associations, the Commissioner must publish a gazetted notice that provides at least 60 days for one to be formed for the purpose of drafting a code, before exercising the power to draft an industry standard.

Industry codes and standards are being implemented through a two-phased approach. Phase 1 focused on Class 1 material, including child sexual exploitation and abuse material, pro-terror material and extreme crime and violence material. Six codes are now in operation, and standards for the two additional sections of industry are being determined. Phase 2 will focus on Class 2 material (material that would be classified R18+ or X18+, including pornography and other high impact material) and remaining Class 1 material (material depicting a small set of fetish practices).

Under the Act, draft codes are prepared by industry associations and then assessed by the regulator. In other jurisdictions, such as the UK for example, the regulator is responsible for drafting codes in consultation with industry. Whether a code is drafted by industry or by the regulator, the complexity of the code-making process increases with the scope of services that are represented within it. In the Australian context, the

---

[13] *Online Safety Act 2021*, section 150.

scope of services covered by the designated internet services sector is an example of where a broad range of services is covered by a single code or standard. Some regulated industries may also not have clearly identified representative associations in Australia.

Basic Online Safety Expectations

The Basic Online Safety Expectations set out the Government's minimum safety expectations of online service providers, establishing a benchmark for online service providers to take proactive steps to protect the Australian community from abusive conduct and harmful content online. While the Basic Online Safety Expectations do not impose a legally enforceable duty on service providers to implement the expectations, it is an essential part of driving transparency and accountability across online services.[14]

The Act provides for the Minister for Communications to determine the Basic Online Safety Expectations for social media services, relevant electronic services and designated internet services by legislative instrument. The current determination, the *Online Safety (Basic Online Safety Expectations) Determination 2022,* includes the core expectations set out in the Act[15] (required by the Act) as well as additional expectations that are determined by the Minister, and examples of reasonable steps that service providers can take to meet the core or additional expectations.

Core expectations for service providers are set out in section 46 of the Act and include:

- take reasonable steps to ensure that end-users are able to use the service in a safe manner
- take reasonable steps to minimise provision of cyber-bullying, adult cyber-abuse, non-consensual intimate images, Class 1 material, and material that promotes, incites, instructs in, or depicts abhorrent violent conduct
- take reasonable steps to prevent access by children to Class 2 material
- ensure the service has clear and readily identifiable mechanisms that enable end-users to report, and make complaints about, certain material provided on the service; and
- ensure the service has clear and readily identifiable mechanisms that enable end-users to report, and make complaints about, breaches of the service's terms of use.

Additional expectations for service providers are determined by the Minister[16] and include:

- proactively minimise the extent to which material or activity on the service is unlawful or harmful
- prevent anonymous accounts from being used to deal with material, or for activity, which is unlawful or harmful
- consult and cooperate with providers of other services to promote the ability of end-users to use all of those services in a safe manner
- have accessible terms of use, policies and procedures in relation to end-user safety, reports and complaints, and standards of conduct; and
- keep records of reports and complaints about certain material.

The Act enables the Commissioner to require service providers to report against the expectations contained in the Basic Online Safety Expectations Determination (through periodic and non-periodic reporting notices and determinations). The reporting requirement aims to boost the transparency of services and provides the Commissioner with a tool to hold services to account for the steps they take to keep Australians safe online.

---

[14] *Online Safety Act 2021*, section 45.
[15] *Online Safety Act 2021*, section 46.
[16] Online Safety (Basic Online Safety Expectations) Determination 2022.

There are no penalties for a service provider failing to comply with the expectations outlined in the Basic Online Safety Expectations Determination. However, a service provider that fails to comply with a reporting notice or determination issued by the Commissioner may be subject to a formal warning or a civil penalty of up to 500 penalty points (currently $782,500) for corporations. Further, if the Commissioner finds that a service has not complied with one or more applicable expectations, the Commissioner can prepare and publish a service provider notification to this effect. The Commissioner may also prepare and publish service provider notifications about a service's failure to comply with a reporting notice. These measures boost transparency for users and create reputational risks for service providers, encouraging improvements to policies, processes, and human and technological interventions to keep Australian end-users safe on their platforms.

In November 2023, the Minister commenced public consultation on a range of amendments to the Basic Online Safety Expectations Determination to address emerging online safety issues, and improve overall operation. Key proposed reforms include inserting new additional expectations that:

- generative artificial intelligence capabilities are designed and implemented with user safety in mind, and that services using generative artificial intelligence capabilities proactively minimise the extent to which that capability produces unlawful or harmful material
- recommender systems are designed and implemented in a manner that enables their safe use, and that services minimise the extent to which recommender systems amplify unlawful or harmful material
- the best interests of the child are a primary consideration in the design and operation of services likely to be accessed by children
- service providers make available controls that give end-users autonomy to support safe online interactions, and
- service providers review and respond to reports and complaints within a reasonable period of time, and provide feedback to users on the actions taken.

Public consultation closed on 16 February 2024, and the Government is considering the outcomes.

Regulated sections of industry

The Act identifies eight sections of the online industry that provide services in Australia:

- Social media services
- Relevant electronic services
- Designated internet services
- Internet search engines
- App distribution services
- Internet carriage services
- Hosting services which host content in Australia
- Manufacturers, suppliers and installers of equipment for use by end-users in Australia in connection with a social media service, relevant electronic service, designated internet service or internet carriage service.

The definitions are based on the primary purpose of the service, such as defining social interactions and the posting of content as 'social media services' and defining messaging between online users as 'relevant electronic services.' However, with industry changing its service offerings, defining the industry in this way may not provide sufficient flexibility for the regulatory framework to adapt to these changes. For example, the increasing convergence between messaging (a relevant electronic service) and social interaction services (a social media service) may blur these industry definitions.

Relevant electronic services include services with highly varied characteristics and risks (such as gaming, dating, and messaging services). Designated internet services are broadly defined to provide the regulator with significant flexibility to capture a wide set of services. However, given the services that fall within this category present very different levels of risk to users, it does add complexity to drafting a single industry code. There is also a risk that the current approach to defining sections of the online industry does not keep pace with the evolving digital environment. For example, with respect to phase 2 codes, the designated internet services category includes both websites providing pornographic material and websites that do not host any adult content. In other jurisdictions, some regulatory approaches are based on the risk and reach of services (such as EU and UK legislation).

Table 2.1 – Summary of service providers regulated through the Basic Online Safety Expectations and Online Content Scheme

Penalty values as at 22 April 2024

Scheme	Industry sectors	Scope	Penalties
Basic Online Safety Expectations	• Social media services • Relevant electronic services • Designated internet services	The Basic Online Safety Expectations includes expectations relating to the safe use of a service, minimising unlawful or harmful material on a service, having identifiable mechanisms to report certain unlawful or harmful material on a service or breaches of a service's terms of use, and enforcing terms of use where there are breaches. The Act provides the Commissioner with powers to require service providers to report on how they are meeting the Expectations.	No penalties for a failure to meet the expectations outlined in the Basic Online Safety Expectations. Penalties of up to 500 penalty points (currently $782,500) for corporations for failure to comply with a periodic or non-periodic reporting notice or determination from the Commissioner.
Online Content Scheme (development of industry codes and standards relating to Class 1 and Class 2 material)	• Social media services • Relevant electronic services • Designated internet services • Internet search engine services • App distribution services • Hosting services • Internet carriage services • Manufacturers, suppliers and installers of equipment	Class 1 material: online material that is a film, publication, computer game or other material that is or would likely to be classified 'Refused Classification'. Class 2 material: online material that is a film, publication, computer game, or other material, that is or would be classified R18+/X18+.	Penalties of up to 500 penalty points (currently $782,500) for corporations for failure to meet a direction from the Commissioner to comply with an industry code, or for failure to comply with an industry standard. Federal Court order to cease providing service may apply where there are two or more civil penalty contraventions in 12 months that create a significant community safety risk.

Table 2.2 - Global comparison of Regulated service providers

Jurisdiction	Regulated Services	Examples of risk-based regulation
United Kingdom	**User-to-user platforms** Where users can upload and share content (for example messages, images, videos, comments) that becomes accessible to others. This includes services such as online discussion forums, social media platforms, dating services and online market places. **Search services** Search engines that enable users to search numerous websites and databases. **Services that provide pornographic content**	Additional regulatory requirements apply to user-to-user platforms with higher reach and risk (Category 1) and highest reach search engines with higher reach (Category 2A) and other services with potentially high-risk functionalities (Category 2B). (The basis for categorisation will be provided in regulations). Specific requirements apply where children are likely to access the service and for services that provide pornographic content.
European Union	**Online Platforms** A hosting service that, at the request of a recipient of the service, stores and disseminates information to the public, unless that activity is a minor and purely ancillary feature of another service or a minor functionality of the principal service and, for objective and technical reasons, cannot be used without that other service, and the integration of the feature or functionality into the other service is not a means to circumvent the applicability of this Regulation. **Online Search Engines** An intermediary service that allows users to input queries in order to perform searches of, in principle, all websites, or all websites in a particular language, on the basis of a query on any subject in the form of a keyword, voice request, phrase or other input, and returns results in any format in which information related to the requested content can be found.	Greatest regulatory requirements apply to very large services that are designated based on regional user base: • **very large online service providers** For example, Facebook, Google Maps, LinkedIn, Instagram, TikTok and YouTube. • **very large online search engines** For example, Bing and Google Search.
Ireland	**Video-sharing platform services** **Other online services designated** by the Media Commission (can range from social media and online gaming to private messaging services).	The Online Safety Commissioner can designate relevant online services (or categories of services) to which codes apply, considering the nature and scale of the service and the levels of risk of exposure to harmful online content when using the service.
Canada (proposed)	**Social Media Services** A website or application that is accessible in Canada, the primary purpose of which is to facilitate interprovincial or international online communication among users of the website or application by enabling them to access and share content (includes adult content services and live streaming services).	Social media services are regulated if they meet a threshold number of users (threshold to be provided in regulations) or are designated in the regulations (if user threshold is not met).

Part 2: Australia's regulatory approach to online services, systems and processes - consultation questions

1. Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?
2. Does the Act capture and define the right sections of the online industry?
3. Does the Act regulate things (such as tools or services) that do not need to be regulated, or fail to regulate things that should be regulated?
4. Should the Act have strengthened and enforceable Basic Online Safety Expectations?
5. Should the Act provide greater flexibility around industry codes, including who can draft codes and the harms that can be addressed? How can the code drafting process be improved?
6. To what extent should online safety be managed through a service provider's terms of use?
7. Should regulatory obligations depend on a service provider's risk or reach?

Part 3 – Protecting those who have experienced or encountered online harms

Overview

> **Harmful online content and behaviour can be seriously damaging, especially for those most at risk. The social, emotional, psychological and even physical impacts of online harms can be immediate, experienced over time and/or enduring.**
>
> World Economic Forum, '*Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms,*' August 2023.

The spectrum of potential online harms is broad, arising in the production, distribution and consumption of online content, as a result of contact with others online, and from harmful conduct or behaviour facilitated by technology or services.[17] No government can completely protect its citizens from online harm. However, where the harm is significant, individuals need appropriate and effective actions to have harmful material removed.

Online abuse can have wide-reaching impacts on a person's life and wellbeing. The online environment amplifies the threat of harm with faster spread and wider audience-reach available.

The Commissioner provides individual support through four complaints and content-based removal notice schemes, and can require the blocking of material depicting abhorrent violent conduct. There is no equivalent to Australia's complaints-based schemes in the EU or UK. In the EU, there are notice and takedown provisions for illegal material, with the EU *Digital Services Act 2022* introducing trusted flagger provisions.

Australia's regulatory framework allows individual complaints while also focusing on systems through the Basic Online Safety Expectations, and the industry codes and standards. Investigating individual complaints can be resource intensive, but content removal schemes can make a significant difference to the targeted individual and limit the harm experienced. The volume of complaints received through the child cyberbullying and adult cyber-abuse schemes, where a person must have complained to the platform before a removal notice can be issued, indicate that eSafety's interventions remain an important protection for Australians.

Complaints and content-based removal notice schemes

There are four complaints and content-based schemes under the Act: the child cyberbullying scheme, the adult cyber-abuse scheme, the non-consensual sharing of intimate images ('image-based abuse') scheme, and the Online Content Scheme. Each focuses on specific types of harmful online material.

The Commissioner has powers to investigate complaints made under these schemes. For each scheme, the Commissioner can issue a removal notice as a formal compliance mechanism, generally to social media services, relevant electronic services, designated internet services, hosting service providers, and in some instances to the individual who posted the harmful material.

The recipient of the removal notice must remove or take all reasonable steps to remove the relevant material or take reasonable steps to cease the hosting of the material within 24 hours (or longer as specified by the

---

[17] World Economic Forum (2023) *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*, August 2023, WEF_Typology_of_Online_Harms_2023.pdf (weforum.org), 5, accessed 26 April 2024.

Commissioner). Failure to comply with a removal notice carries a civil penalty of up to 500 penalty points ($782,500 as at 22 April 2024) for corporations.

Child cyberbullying scheme

The Act expanded Australia's world-first cyberbullying scheme to provide protection to children being bullied in all online environments, not just on social media.

The Act enables complaints to be made to eSafety about child cyberbullying material that is targeted at a child who is ordinarily resident in Australia. Complaints may be made by the child, a responsible person on behalf of the child, or an adult who was a child if a complaint was made within six months after the person reached 18 years.[18] 'Cyberbullying material' is defined under the Act to refer to material that 'would be likely to have the effect of . . . seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child'.[19]

In the 2022-23 financial year eSafety received 1,969 complaints in relation to child cyberbullying material:[20]

- 636 informal requests were made to remove child cyberbullying material, with an 84 per cent success rate in having the material removed.
- 13 formal end-user notices were issued requiring individuals to remove child cyberbullying material and cease cyberbullying the target.

eSafety finds that in some circumstances it is more appropriate to take informal action, such as contacting the online service or the individual directly. This will often result in a more expeditious removal of the harmful material as it does not involve the preparation of a formal notice and can be preferred where children are directly involved.

Further information on the cyberbullying scheme can be found here.

Adult cyber-abuse scheme

The Act introduced a world first adult cyber-abuse scheme for Australians. It enables complaints to be made to the Commissioner about cyber-abuse material that is targeted at an adult ordinarily resident in Australia. A complaint may be made by the targeted adult or a responsible person authorised by the targeted adult if the material is posted on a social media service, designated internet service or relevant electronic service. A service is exempt if the material on the service is not accessible to, or delivered to, an end-user in Australia.[21]

'Cyber-abuse material' is defined to mean material that:

- an ordinary reasonable person would conclude was likely intended to cause serious harm to an Australian adult; and
- an ordinary reasonable person in the position of the targeted Australian adult would regard as being, in all the circumstances, menacing, harassing or offensive.[22]

The threshold for regulatory action is extremely high for adult cyber-abuse, and higher than for child cyberbullying to reflect that adults generally have a higher level of resilience than children, and to ensure that freedom of expression is not unduly restricted. That means that if a post (or collection of posts) is offensive

---

[18] *Online Safety Act 2021*, section 30.
[19] As concluded by a reasonable ordinary person. *Online Safety Act 2021*, section 6.
[20] eSafety Annual Report 2022-23, available here.
[21] *Online Safety Act 2021*, sections 13-14.
[22] *Online Safety Act 2021*, section 7.

but would not be considered by an ordinary person to be likely intended to cause serious harm, it would not be considered adult cyber-abuse material under the Act.

The term 'adult cyber-abuse' in the Act is currently reserved for the most severely abusive material intended to cause serious psychological harm or serious physical harm. This includes material which sets out realistic threats, places people in real danger, is excessively malicious or is unrelenting. That means the Commissioner may be unable to require removal of abusive online material targeted at an Australian adult, including adults in vulnerable or high-risk situations, because the material does not meet the criteria of adult cyber-abuse under the Act. Examples of situations where the Commissioner may not be able to intervene include:

- a person affected by abusive posts targeted at a group of people (such as dehumanising commentary on a particular race or religious belief)
- a person impacted by abusive posts targeted at another person (such as support staff for public figures)
- a person unable to remove themselves from exposure to persistent abuse because their position requires them to have an online presence
- a person experiencing financial or personal harm from attacks directed at their business
- a person experiencing volumetric or pile-on attacks, where individual posts do not meet the definition of adult cyber-abuse
- a person ordinarily resident overseas who is targeted by an Australian.

In the 2022-23 financial year 2,516 complaints were made relating to adult cyber-abuse. Of these, 877 related to reputational harm which in almost all instances did not meet the threshold of adult cyber-abuse. For these complaints, eSafety provides guidance about self-reporting to the platform, resources for obtaining legal advice, general resources about the difference between cyber abuse and defamation, and information about obtaining support services.

Of those complaints that did meet the threshold, eSafety informally approaches platforms about removal in the first instance. If that is unsuccessful, a formal notice is issued where possible.

Of those complaints that did meet the threshold:

- the Commissioner issued three removal notices, with material removed in all three cases.
- eSafety made 601 informal requests to online service providers seeking removal of material for matters related to the terms of service, with the material successfully removed for 466 of these.

Further information on the adult cyber-abuse scheme can be found here.

Non-consensual sharing of intimate images scheme (Image-based abuse)

The Act updated Australia's non-consensual sharing of intimate images scheme[23] to address the sharing and threatened sharing of intimate images without the consent of the person shown. It enables complaints to be made to the Commissioner for the posting or sharing of intimate images of another person without their consent. The scheme applies to images posted on a social media service, a relevant electronic service, or a designated internet service. Recently, the scheme saw a significant rise in the number of reports relating to sexual extortion (sextortion), with the majority of the reports from young men aged between 18 and 24.

Complaints or objections may be made by a person who has reason to believe they are depicted in the image, an authorised person on behalf of the depicted person, or a parent or guardian of the depicted person (if that person is a child who has not reached 16 years, or is temporarily or permanently incapable of managing their

---

[23] Formerly provided for in the *Enhancing Online Safety Act 2015.*

own affairs). The image must also have a link to Australia (either depicting a person that is ordinarily resident in Australia or posted by an end-user ordinarily resident in Australia). An objection can also be made if the depicted person previously consented to the provision of the intimate image on the service and the image is hosted in Australia by a hosting service.

The images or videos must show the person in circumstances where they would reasonably expect to be afforded privacy. The Act specifies that it is not relevant if the material has been altered in any way (so complaints or objections can be made about images that have been digitally altered, including deepfakes and photoshopped images).[24]

In the 2022-23 financial year, the Commissioner received 9,060 complaints in relation to image-based abuse:

- The Commissioner requested the removal of intimate images from more than 6,500 locations (generally URLs) across 340 platforms and services, primarily on pornography websites hosted overseas. 87 per cent of this material was removed through informal removal requests.
- 15 removal notices were issued. In 14 matters, content was either entirely or partially removed. In the 15th matter, the content was not removed but discoverability was restricted.
- Two infringement notices were issued which are part of an ongoing matter currently before the Federal Court.

Further information on the image-based abuse scheme can be found here.

Online Content Scheme

The Online Content Scheme addresses the accessibility of illegal and restricted material online for end-users in Australia. It regulates Class 1 and Class 2 material that is provided by various sections of the online industry. [25]

The Act updated Australia's existing Online Content Scheme,[26] providing new powers to regulate illegal and restricted content, no matter where it is hosted.[27] It enables the Commissioner to take action if they have reason to believe end-users in Australia can access either Class 1 or Class 2 material; this could be based on a complaint.[28] A complaint can be made by an individual who resides in Australia, a body corporate that carries on activities in Australia, or by the Commonwealth or a State or Territory.

'Class 1' and 'Class 2' materials are defined by reference to the National Classification Code. 'Class 1' material is material that is, or would likely be, refused classification (such as child sexual exploitation and abuse material and pro-terror material).[29] 'Class 2' material is material that is, or would likely be, classified R 18+ or X 18+ (such as pornography, and other high impact material).[30] Under the Act, in the exercise of certain powers the Commissioner must be satisfied that material is either Class 1 or Class 2 material. The Act also permits the Commissioner to obtain advice from the Classification Board about whether particular material is Class 1 or Class 2 material at any time.

---

[24] *Online Safety Act 2021*, subsection 15(5).
[25] Social media service, relevant electronic service, designated internet service or hosting service (in relation to removal notices).
[26] Previously provided for under Schedule 5 and Schedule 7 of the *Broadcasting Services Act 1992*.
[27] A removal notice for Class 1 material may be issued no matter where the content is hosted. For Class 2 material, the service must be provided from Australia.
[28] *Online Safety Act 2021*, section 38.
[29] *Online Safety Act 2021*, section 106.
[30] *Online Safety Act 2021*, section 107.

The Government is separately progressing reforms to ensure the National Classification Scheme is fit-for-purpose in the modern media environment. Public consultation on options for second stage reforms commenced with the release of a public consultation paper on 4 April 2024. Further information about that public consultation process can be found here.

eSafety is the Australian member of the International Association of Internet Hotlines (INHOPE), which allows for the referral of child sexual exploitation and abuse material between network members and relevant law enforcement agencies for rapid removal in the country where it is hosted. If material is hosted in a non-INHOPE country, eSafety informs the Australian Federal Police and may seek removal action under the Act.

In the 2022-23 financial year, the Commissioner received 11,636 complaints about 33,122 URLs in relation to illegal and restricted content:

- 87 per cent related to child sexual exploitation and abuse, child abuse or paedophile activity
- 14,975 notifications were sent to the INHOPE Network and there were 76 referrals to the Australian Federal Police
- The Commissioner issued three formal removal notices, and two link deletion notices to an internet search engine provider.

Due to the nature of the material being reported, all reports about illegal and restricted online content can be made anonymously. eSafety accepts anonymous reports to encourage reporting. Most complaints received about child sexual exploitation and abuse material are anonymous. To make a complaint online, the complainant must be an Australian resident, this includes an individual, a business, an organisation or other government agencies, including law enforcement.

Child access to pornography and violent pornographic content

Pornographic content is regulated by the Online Content Scheme.

The Act provides for age assurance through the Online Content Scheme, the Restricted Access System and the Basic Online Safety Expectations. Under the Online Content Scheme, phase 2 industry codes and/or standards will be developed to focus on Class 2 material, which includes pornography and other high impact online material that would be classified R18+ or X18+ under Australia's National Classification Code.

The eSafety Commissioner also has the power to declare an access-control system is a restricted access system for the purposes of the Act. The current *Online Safety (Restricted Access Systems) Declaration 2021* seeks to ensure that the methods for limiting access to relevant Class 2 material meet a minimum standard. In certain circumstances, the Commissioner can investigate access to Class 2 material by end-users in Australia, investigate any such material hosted in Australia or publish a statement that certain Class 2 material was not subject to a Restricted Access System.

Children encountering pornography and Australians encountering violent pornographic content online remain areas of community concern.[31] Unintentional and deliberate encounters with pornography are widespread

---

[31] Maree Crabbe, Michael Flood and Kelsey Adams (2024) Pornography exposure and access among young Australians: a cross-sectional study, *Australian and New Zealand Journal of Public Health,* 1. Crabbe, Flood, Adams found there is some evidence suggesting that young people's exposure to pornography may have public health implications, in particular in 'shaping young people's sexual understandings, expectations, and experiences.'

among young people,[32] with some evidence suggesting that young people's exposure to pornography may have public health implications, in particular in 'shaping young people's sexual understandings, expectations, and experiences.'

Some forms of pornography have been found to include significant levels of violent, sexist, and racist content, with aggression overwhelmingly directed toward women and typically by men.[33] In Australia, a 2022 report showed that 23 per cent of young people aged 14-17 had encountered violent sexual images or videos online.[34] Pornography often fails to depict relational intimacy, safe sex, or the negotiation of consent.[35] The consumption of pornography has been associated with a range of harmful attitudes, behaviours and experiences including risky sexual practices, acts of sexual aggression, sexual violence, stronger beliefs in gender stereotypes, and more sexualised and sexually objectifying views of women.[36] However, these associations do not demonstrate causation and further longitudinal studies are required to clarify the directionality of these effects.

Although both young men and young women see pornography, evidence suggests that young men are more likely to encounter online pornography at a younger age and more frequently.[37] For example, a recent Australian study found that young men are encountering online pornography more frequently than young women, with 21 per cent of young men encountering this content daily compared to 4 per cent of young women.[38] 2018 findings from the Longitudinal Study of Australian Children indicated that more frequent encounters with pornography by young men was associated with greater likelihood of engaging in unwanted sexual behaviours (such as showing or sending sexual pictures, stories or jokes that made someone feel uncomfortable; making sexual gestures, rude remarks, touching, or looking at someone in a way that embarrassed or upset them; and repeatedly asking someone out on a date, or asking them to hook up, although they said 'No').[39]

Possible approaches to limiting child access to pornography were raised in 2023 in the Commissioner's Roadmap for Age Verification and complementary measures to prevent and mitigate harms to children from online pornography.[40] The Government is currently scoping an age assurance pilot, working across departments, given the need for cross-portfolio engagement on this issue. This pilot would complement the development of phase 2 codes. The scoping work is having regard to international and industry developments.

---

[32] eSafety Commissioner (2023), *Accidental, unsolicited and in your face. Young people's encounters with online pornography: a matter of platform responsibility, education and choice,* Accidental-unsolicited-and-in-your-face.pdf (esafety.gov.au), accessed 19 April 2024.

[33] eSafety Commissioner (2023), *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography*, August 2023, Age-verification-background-report.pdf (esafety.gov.au), 76-77, accessed 26 April 2024.

[34] eSafety Commissioner (2022), *Mind the Gap: Parental awareness of children's exposure to risks online*, August 2022, Mind the Gap – Parental awareness of children's exposure to risks online, accessed 26 April 2024.

[35] Maree Crabbe and Michael Flood (2021) School-based education to address pornography's influence on young people: A proposed practice framework, *American Journal of Sexuality Education*, 16(1), 1-37.

[36] eSafety Commissioner (2023), *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography*, August 2023, Age-verification-background-report.pdf (esafety.gov.au), accessed 26 April 2024.

[37] eSafety Commissioner (2023). *Accidental, unsolicited and in your face. Young people's encounters with online pornography: a matter of platform responsibility, education and choice*. Accidental-unsolicited-and-in-your-face.pdf (esafety.gov.au), accessed 26 April 2024.

[38] eSafety Commissioner (2023). *Accidental, unsolicited and in your face. Young people's encounters with online pornography: a matter of platform responsibility, education and choice*, Accidental-unsolicited-and-in-your-face.pdf (esafety.gov.au), accessed 26 April 2024.

[39] Diana Warren and Neha Swami (2024), *Teenagers and sex*, Australian Institute of Family Studies LSAC Annual Statistical Report 2018 Chapter 5 Teenagers and sex (growingupinaustralia.gov.au), accessed 26 April 2024.

[40] eSafety Commissioner (2023), *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography,* March 2023, 29 (Recommendation 2), Roadmap-for-age-verification_2.pdf (esafety.gov.au), accessed 26 April 2024.

Posting harmful or criminal material to increase notoriety

The Act provides the Commissioner with powers to require the removal of Class 1 material that would be refused classification in Australia, including material that promotes, incites or instructs people in crime or violence. The Commissioner can also use their well-established relationships with law enforcement and social media platforms, to have material removed expeditiously.

However, concerns remain in the community about criminal or violent activity being posted on social media to increase 'likes'. In March 2023, the Queensland Government introduced new penalties for people who boast about crime on social media. In March 2024, the New South Wales Government announced the introduction of law reforms to include a new offence for people who commit, then 'post and boast' about motor vehicle theft and break and enter offences.[41] These issues have also been raised in the Australian Parliament as recently as March 2024 through private members bills and other parliamentary business.

Operation of the content and complaints-based removal schemes

Over time the Act has been expanded to bring on new complaints-based schemes, including a new adult cyber-abuse scheme. There are variations across the schemes that may add unnecessary regulatory complexity, for both the Commissioner and those seeking to make a complaint. Tables 3.1 and 3.2 outline these variations, including who can report, who is protected, the link required to Australia, and available regulatory actions.

For all complaints-based schemes, except the Online Content Scheme, complaints must be made by the individual targeted or depicted by the material. The only exceptions to this are where a complaint is made by someone authorised by the individual or by the parent or guardian of the individual, where the individual is a child or is mentally or physically incapacitated. These conditions create some risk that harmful material targeting or depicting Australians may spread before a targeted individual becomes aware of it, which may be particularly problematic in the case of image-based abuse (such as deepfake intimate images). A 'bystander', or member of the general public, is not currently able to report such material to eSafety, even in circumstances where it may be reasonable to suppose the target would be harmed by it, and would not have consented to it being posted.

For child cyberbullying and adult cyber-abuse complaints, the complainant must report to the online platform first in order for eSafety to give a removal notice. Requiring complainants to report the material directly to an online service can carry its own risks, particularly where that service might be motivated to act against the complainant with malice. This is particularly common in the case of websites set up to 'dox' complainants. The Commissioner may only issue a formal removal notice if the platform has not removed the material within 48 hours of the complaint, providing a window in which the online harm can amplify.

Of all the complaints schemes in the Act, the Online Content Scheme has the broadest scope in terms of who can make a complaint, the services regulated, and the basis for making a complaint. The other three schemes (cyberbullying, adult cyber-abuse and image-based abuse) are limited to complaints from targeted individuals or their representatives. A complaint can be made by a person or government based in Australia if it is suspected Australians can access Class 1 or Class 2 material (illegal and restricted online content).

The Australian connection to the Online Content Scheme is broad and based on the ability of end-users in Australia to access illegal or restricted material. However, the child cyberbullying and adult cyber-abuse schemes require the targeted individual to be ordinarily resident in Australia. The image-based abuse scheme applies to both end-users ordinarily resident in Australia who either post or who are targeted online, and

---

[41] NSW Government (2024) NSW Government takes action to make communities safer and support young people in regions, 12 March, NSW Government takes action to make communities safer and support young people in regions | NSW Government, accessed 26 April 2024.

extends to images hosted in Australia (for objection notices only). Powers to address the most harmful (Class 1) material do not require the material to be hosted in Australia.

While children are widely recognised as among the most at-risk in relation to online harms, other groups of Australians also have a greater risk of abuse online. The risk of online abuse is greater for women (and women in public or prominent positions), people from culturally and linguistically diverse backgrounds, people living with disability or medical conditions, Aboriginal and Torres Strait Islander peoples, people who identify as LGBTQIA+, people with particular religious beliefs, and older Australians.[42]

Table 3.1 – Overview of complaint and content-based removal schemes

	Image-Based Abuse	Child Cyberbullying	Adult Cyber-Abuse	Illegal and Restricted Content
Online harm	Posting or threatening to post an intimate image depicting another person without that person's consent (irrespective of whether the image has been altered).	Online material that is likely intended to have an effect on an Australian child, and likely to have the effect of seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child.	Online material that is likely intended to have the effect of causing serious harm to an Australian adult and would reasonably be regarded as menacing, harassing or offensive in all the circumstances.	Online material that is Class 1 or Class 2 material (determined by reference to Australia's National Classification Code).
Who is protected?	Person depicted (or purported to be depicted).	A targeted child (who is ordinarily resident in Australia).	A targeted adult who is ordinarily resident in Australia.	End-users in Australia.
Link required to Australia	The person depicted, or the person who posted or threatened to post, is ordinarily resident in Australia (or, for objection notices only, the image is hosted in Australia).	Material is targeted at a child ordinarily resident in Australia ('Australian child').	Material is targeted at an adult ordinarily resident in Australia ('Australian adult').	Material suspected to be accessible to Australians online. Class 1 material can be hosted anywhere, but Class 2 material must be provided by a service in Australia or hosted in Australia.
Who can make a complaint?	The person who has reason to believe an intimate image depicting them has been shared without consent (or that a threat to share such an image has been made); a person authorised by the depicted person; or a parent or guardian of the depicted person.	The targeted Australian child or a parent, guardian or responsible person authorised by the child or an adult who was an Australian child.	The targeted Australian adult or responsible person authorised by the Australian adult.	A person who resides in Australia, or an entity that carries out activities in Australia, or an Australian Government. (Note, eSafety can investigate material within this scheme without receiving a complaint).
Does complainant need to report to the service provider before a removal notice can be issued?	No.	Yes.	Yes.	No (can be reported anonymously).

---

[42] Parliament of the Commonwealth of Australia, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' March 2022, [2.88]-[2.113], Social Media and Online Safety – Parliament of Australia (aph.gov.au), accessed 26 April 2024.

Table 3.2 – Commissioner's complaints scheme compliance and enforcement powers

	Image-Based Abuse	Child Cyberbullying	Adult Cyber-Abuse	Online Content Scheme
Formal warning to person who posts or threatens to post image	Yes	No	No	No
Removal notice to service provider/hosting service provider	Yes	Yes	Yes	Yes
Removal notice to end-user	Yes (a 'removal notice')	Yes	Yes (a 'removal notice')	No
Remedial direction to end-user	Yes	No	No	No
Remedial notice to service provider	No	No	No	Yes (Class 2 only)
Service provider notification	Yes	Yes	Yes	No
Service provider statement	Yes	Yes	Yes	Yes
Link deletion notice	No	No	No	Yes (Class 1 only)
App removal notice	No	No	No	Yes (Class 1 only)
Federal Court order to cease providing service	No	No	No	Yes (in exceptional situations)
Alternative enforcement arrangements	Formal warnings, enforceable undertakings, court injunctions, infringement notices, civil penalty orders and financial penalties.			

These compliance and enforcement powers involve the following:

- **Removal notice:** Notice requiring recipient (end-user, service provider or hosting service provider) to remove material or stop hosting material within 24 hours or longer period the Commissioner allows (civil penalty for non-compliance).
- **End-user notice:** Notice requiring person who posted cyberbullying material targeted at a child ordinarily resident in Australia to: remove the material and/or refrain from posting cyber-bullying material targeting the child and/or apologise for posting the material (enforceable by injunction).
- **Remedial direction:** Direction to end-user who has posted or threatened to post intimate images without consent to take specified remedial action to prevent future contraventions (civil penalty for non-compliance).
- **Remedial notice:** Notice requiring the recipient to remove Class 2 material or to make the material subject to a Restricted Access System (civil penalty for non-compliance).[43]

---

[43] The Commissioner may declare by legislative instrument that a specified access control system is a restricted access system (*Online Safety Act 2021*, section 108).

- **Link deletion notice:** Notice requiring internet search engine provider to cease providing a link to Class 1 material where material subject to a removal notice in the last 12 months has not been removed and the link has been used to access the material at least twice in a 12-month period (civil penalty).
- **App removal notice:** Notice requiring an app distribution service to cease the ability for end-users in Australia to download an app used to facilitate the posting of Class 1 material at least twice in a 12-month period and where the material has not been removed following a removal notice issued in the past 12 months (civil penalty).
- **Service provider notification:** Notification to service provider advising that material on the service has been found to be the specific harmful material defined under the complaint scheme.
- **Service provider statement:** A statement that the Commissioner has found multiple occurrences of harmful material being posted/hosted on the service within a 12-month period, in contravention of the service's terms of use, and that may be published on the Commissioner's website.
- **Federal Court order:** In the most exceptional situations, the Commissioner can apply for Federal Court orders for a social media service, relevant electronic service, designated internet service or internet carriage service to cease providing their service in Australia. An order can only be made where a service provider has, on two or more occasions in the past 12 months, contravened a civil penalty provision in the Online Content Scheme and as a result the continued operation of the service represents a significant community safety risk.
- **Alternative enforcement arrangements**: The Act adopts the enforcement arrangements set out in the *Regulatory Powers (Standard Provisions) Act 2014* for civil penalties, infringement notices, enforceable undertakings and injunctions. The Commissioner can issue formal warnings, issue an infringement notice, accept an enforceable undertaking, seek a court ordered injunction, and/or pursue civil penalties for non-compliance with a requirement under the Act depending on the case circumstances. eSafety's *Compliance and Enforcement Policy* outlines the matters considered when determining what the preferred compliance and enforcement actions are in a particular situation.

Material that depicts abhorrent violent conduct

In addition to the complaints-based removal schemes, the Act also provides a mechanism for the Commissioner to request or require the blocking of material that promotes, incites, instructs in or depicts abhorrent violent conduct if the material is likely to cause significant harm to the Australian community. 'Abhorrent violent conduct' is defined under the Criminal Code and occurs when a person engages in particularly egregious conduct, including engaging in a terrorist act, murdering or attempting to murder another person, or torturing another person.[44]

'Material that depicts abhorrent violent conduct' is defined in section 9 of the Act, and includes audio material, video material, or audio-visual material that records or streams abhorrent violent conduct. It is immaterial whether the material has been modified, or who has produced it. The online spread of such material, as evidenced through the livestreamed attack by a gunman in Christchurch in 2019, can have a seriously harmful impact on the Australian community.

Examples of steps the Commissioner could require in a blocking notice or request include blocking domain names, URLs or IP addresses that provide access to the material. Each blocking notice or request can apply for up to three months. If a notice is about to expire but the material still needs to be blocked, the Commissioner can issue a new blocking notice that comes into force immediately after the expiry of the original notice.

---

[44] Criminal Code, subsection 474.32(1).

The Act's provisions for dealing with material depicting abhorrent violent conduct are distinct from the provisions of the Criminal Code to deal with abhorrent violent material. Criminal Code provisions apply only to material produced by the perpetrator or an accomplice.[45] Under the Criminal Code, the eSafety Commissioner may issue notices to services to inform them they are hosting abhorrent violent material. Notices do not require the material to be removed. However, if a service is later prosecuted for failing to remove or cease hosting the material, the notice can be used in legal proceedings to show recklessness.

The Basic Online Safety Expectations also set out a core expectation that service providers will take reasonable steps to minimise the extent to which material that promotes, incites, instructs in or depicts abhorrent violent conduct is provided on the service, and that the service has clear and readily identifiable mechanisms that enable end-users to report, and make complaints about such material.[46]

Harm prevention: Online safety education and promotion

There is an important role for Governments in creating the right policy settings for harm prevention and harm mitigation, but there are also actions individuals can take to keep themselves safe online.

The volume and distribution of potentially harmful material online, and the potential for rapid spread and escalation, mean that educating and empowering Australians online are key elements of an online safety regulatory framework. The Commissioner has a strong focus on harm prevention through education and outreach activities to try to lessen the likelihood of online harms occurring. eSafety also has a research and evaluation function to ensure programs, policies and regulatory functions are informed by evidence.

Under Section 27 of the Act, it is the Commissioner's role to:

- promote online safety for all Australians
- coordinate activities of Commonwealth Departments, authorities and agencies relating to online safety for Australians
- support and encourage the implementation of measures to improve online safety for Australians
- collect, analyse, interpret and disseminate information relating to online safety for Australians
- support, encourage, conduct, accredit and evaluate educational, promotional and community awareness programs that are relevant to online safety for Australians
- make grants of financial assistance in relation to online safety for Australians
- support, encourage, conduct and evaluate research about online safety for Australians.

Online safety education

eSafety undertakes a range of education activities to help inform Australians about how to stay safe online and use technology safety. These include promoting best practice standards, developing education resources and delivering training sessions and webinars.

Training is provided to key audiences with direct influence on children and young people, including educators, parents and carers. Other audiences include frontline workers supporting people experiencing family, domestic and sexual violence, those working with clients in vulnerable situations and communities, senior Australians, and others. An evidence-based co-design process is used to develop resources, training and education materials. This draws on experiences from within the relevant cohort, to ensure the resources developed will be fit for purpose.

---

[45] Criminal Code, subsection 474.31(c): a person who engaged in, conspired to; aided, abetted, counselled or procured, or was in any way knowingly concerned in; or attempted to engage in the abhorrent violent conduct.

[46] *Online Safety Act 2021*, section 46.

eSafety has developed a Best Practice Framework for Online Safety Education. This aims to establish a consistent national approach to online safety education that supports education systems across Australia to deliver high quality programs with clearly defined elements and effective practices.

Be Connected training courses are aimed at developing the digital skills of older Australians so they can confidently engage with digital devices and use the internet safely. In 2022-23, there were 3,171 attendees at Be Connected presentations and webinars.[47]

In April 2022, based on recommendations from youth engagement and online safety research, eSafety set up the Youth Council, made up of members aged 13-24 from diverse locations, genders and backgrounds. The Youth Council makes sure that that young people's views and experiences are considered when developing resources, determining priority areas, and improving engagement and awareness of eSafety among young people.[48]

Professional development and training opportunities are available to frontline workers, such as disability support workers and those who work with people experiencing technology-facilitated abuse. eSafety Women educates frontline workers and specialists about gender-based online violence against women and provides tools for identifying and responding to technology-facilitated abuse. In 2022-23, over 21,000 individuals participated in frontline training and professional learning sessions.[49]

eSafety engages with online safety education providers through the Trusted eSafety Provider program, where approved online safety education providers help raise awareness of eSafety's role and resources when delivering their online safety education programs. In 2022-23, nearly 1.4 million people (including over 1.1 million school students, 140,000 parents and 31,000 educators) participated in training run by education providers endorsed under the Trusted eSafety Provider program.[50]

eSafety provides grants to organisations working to improve online experiences for Australians. In 2022-23, the Online Safety Grants Program provided funding of $2.25 million to recipients under the Trusted eSafety Provider program.[51] eSafety also administers the Preventing Tech-based Abuse of Women Grants Program, which forms part of the Government's commitment to the aims and objectives of the National Plan to End Violence against Women and Children 2022-32. $10 million in funds is available over three years to support initiatives that aim to address or prevent tech-based abuse against women and children.

Online safety promotion

eSafety engages broadly in Australia and internationally to promote online safety. This includes producing educational website content and digital resources, presenting and providing awareness training to stakeholders and a variety of organisations, promoting Safety by Design, conducting and publishing research on online safety and emerging online harms, and collaborating locally and internationally to raise awareness about online safety issues.

eSafety works with industry and technology companies to promote the importance of embedding safety by design in the design, development, and deployment of products and services in an effort to prevent online harms. While this work is gaining traction, the Act has no enforceable requirement to ensure that safety is at the centre of thinking as new technologies and services are developed. It is a core expectation of the Basic Online Safety Expectations 'that the provider of the service will take reasonable steps to ensure that

---

[47] eSafety Annual Report 2022-23, available here.
[48] eSafety Commissioner and Western Sydney University (2022),'*Consultations with Young People to Inform the eSafety Commissioner's Engagement Strategy for Young People*, Young and Resilient Research Centre, available here.
[49] eSafety Annual Report 2022-23, available here.
[50] eSafety Annual Report 2022-23, available here.
[51] eSafety Annual Report 2022-23, available here.

end-users are able to use the service in a safe manner,' however this does not create a legally enforceable duty.

eSafety works with state and territory police, the Australian Federal Police, and the Australian Centre to Counter Child Exploitation to raise awareness about the online safety regulatory schemes and support the work of these agencies. eSafety also established a National Online Safety Education Council to foster cooperation with government, Catholic, and independent school education bodies in each state and territory to help raise awareness of the support and resources eSafety offers, and improve the uptake of the Best Practice Framework for Online Safety Education.

Online safety issues and the online environment are constantly evolving. In 2022-23, eSafety published research on a range of issues including:

- Online experiences of Aboriginal and Torres Strait Islander children and their parents and caregivers
- Australians' negative online experiences 2022
- How adults with intellectual disability experience online abuse
- Risks and benefits of online gaming for children and young people
- Experiences in the metaverse
- The digital experiences of young people with disability
- Young peoples' attitudes towards online pornography and age assurance, and
- Tech-based family, domestic and sexual violence.

eSafety also regularly undertakes evaluation of its education programs, awareness-raising efforts, and regulatory activities to promote accountability, support continuous improvement and innovation, and champion evaluation of online safety initiatives to strengthen the evidence base on what works to prevent and remediate online harms. [52]

In 2022, the Global Online Safety Regulators Network was established, which the Commissioner Chaired in 2022-23.[53] Current members include eSafety, Arcom (France), Coimisiún na Meán (Ireland), Film and Publication Board (South Africa), Korean Communications Standards Commission, Office of Communications (United Kingdom), and the Online Safety Commission (Fiji). Other regulators also participate as observers.

While community engagement with eSafety is increasing, there are still opportunities to increase Australians' awareness of the support eSafety provides. A 2022 National Online Safety Survey identified a parental education opportunity about reporting mechanisms available through eSafety for negative online experiences. Parents have the most prominent role in influencing, monitoring, preventing and intervening in children's online experiences, but reported negative online experiences to eSafety less than children, and only 2.12 per cent mentioned the eSafety Commissioner without prompting as an organisation they would trust to help (this increased to 45.14 per cent when prompted with a list of organisations). The survey also identified opportunities for teachers, carers and supervisors to benefit from training in applying various internet function controls, and for teachers and parents to benefit from receiving additional tips and advice on how to have discussions about online safety with children.[54]

---

[52] In 2022-23, eSafety published evaluations on the Dedicated Project Officer grants program, the Disability workforce and frontline worker program, and the teacher professional learning program.

[53] The purpose of the Global Online Safety Regulators Network is to bring together independent online safety regulators to cooperate across jurisdictions by sharing information, best practice, experience and expertise, and to support harmonised or coordinated approaches to online safety issues. The Terms of Reference for the Network can be found here.

[54] Social Research Centre (2022), *The 2022 National Online Safety Survey – summary report*, 2022 National Online Safety Survey - 7 July 2022 (infrastructure.gov.au), executive summary, accessed 26 April 2024.

Part 3 – Protecting those who have experienced or encountered online harms – consultation questions

8. Are the thresholds that are set for each complaints scheme appropriate?
9. Are the complaints schemes accessible, easy to understand and effective for complainants?
10. Does more need to be done to make sure vulnerable Australians at the highest risk of abuse have access to corrective action through the Act?
11. Does the Commissioner have the right powers to address access to violent pornography?
12. What role should the Act play in helping to restrict children's access to age inappropriate content (including through the application of age assurance)?
13. Does the Commissioner have sufficient powers to address social media posts that boast about crimes or is something more needed?
14. Should the Act empower 'bystanders', or members of the general public who may not be directly affected by illegal or seriously harmful material, to report this material to the Commissioner?
15. Does the Commissioner have sufficient powers to address harmful material that depicts abhorrent violent conduct? Other than blocking access, what measures could eSafety take to reduce access to this material?
16. What more could be done to promote the safety of Australians online, including through research, educational resources and awareness raising?

Part 4 – Penalties, and investigation and information gathering powers

Penalties and enforcement

In Australia, penalties under the Act generally focus on financial penalties directed at individuals or platforms. In most cases, the Act follows the arrangements set out in the *Regulatory Powers (Standard Provisions) Act 2014*, with provisions for civil penalties, enforceable undertakings, infringement notices, and injunctions. Civil penalty provisions in the Act (such as failure to comply with removal notices, industry codes/standards or Basic Online Safety Expectations reporting notices) carry financial penalties of up to 500 penalty units**.** As at 22 April 2024 this means a maximum penalty of $156,500 for individuals and $782,500 for corporations. However, where these provisions require action within a specified period or before a particular time (such as the removal of offending material within 24 hours of a notice being issued), a civil penalty may be applied for *each day* in which the action has not been performed. This may potentially lead to higher penalties.

Broadly speaking however, Australia's penalties regime has not kept pace with newer regulatory regimes, such as in Ireland, the EU, and the UK, which apply significantly higher penalties, including penalties based on a percentage of a platform's global revenue (see Table 4.1 for comparisons). Penalties under the Act are also comparatively low in the Australian context, with potential penalties for platforms under other regimes (such as the Australian Consumer Law and *Privacy Act 1988*) being significantly higher (see Table 4.2 for comparison).

Penalties under the Act may also fail to strike a proper balance between the various offences within the Act itself. For example, the maximum penalty for failing to take down illegal material such as child sexual exploitation material or pro-terror material is the same as for failure to take down harmful but not unlawful material (such as child cyberbullying or adult cyber-abuse material). Penalties under the Act also make no distinction between non-compliance in specific cases (such as the failure to take down a particular offending post) and more systemic non-compliance (such as failure to comply with a code or standard governing Class 1 material in general).

Another issue facing Australia's existing penalties regime under the Act relates to the enforceability of penalties upon individuals or platforms based overseas. While section 23 of the Act formally extends its enforceability to 'acts, omissions, matters and things outside Australia', there can be practical challenges to enforcement outside of Australia. This is a potential challenge for Australia's online safety framework, as the majority of online platforms Australians use and rely on are based overseas, at times with little or no local presence. While this is not an easy issue to resolve, international developments may provide examples of other potential solutions. For example, the UK has established new powers that could be used to stop other companies working with a platform to prevent it from generating money. Options like this could be considered in exceptional circumstances to help address severe non-compliance by disrupting an online services' ability to generate revenue in Australia.

Table 4.1 International comparison of online safety penalties

	Australia *Online Safety Act 2021*	Ireland *Online Safety and Media Regulation Act 2022*	EU *Digital Services Act 2022*	UK *Online Safety Act 2023*	Canada Bill C-63 (Online Harms Act)
Maximum Fines / Penalties	**Civil penalty** for non-compliance with regulatory requirements under the Act of up to: • 500 penalty units ($156,500 for individuals or $782,500 for corporations as at 22 April 2024) A person who contravenes a civil penalty provision that requires an act or thing to be done within a particular period or before a particular time commits a separate contravention of that provision for each day the contravention occurs. The Commissioner may apply for a Federal Court order that a person stop providing a social media service, relevant electronic service, designated internet service, or internet carriage service if the person has contravened a civil penalty provision of the Online Content Scheme (Part 9 of the Act) two or more times in the previous 12 months and as a result the continued service operation represents a significant community safety risk.	**Financial sanctions** of up to: • €20 million or • 10 per cent of annual turnover in the prior financial year attributable to the service that gave rise to the contravention. Regulator can compel a non-compliant service to take certain actions. Regulator can block access to the non-compliant service in Ireland.	**Financial sanctions** Member states are responsible for investigations and setting infringement penalties, up to a maximum value of: • 6 per cent of the intermediary service provider's annual worldwide turnover in the preceding financial year. • Or for periodic penalties, 5 per cent of the average daily worldwide turnover of the intermediary service provider in the preceding financial year per day. As a last resort, if the infringement persists and causes serious harm to users and entails criminal offences involving threat to persons' life or safety, Digital Services Coordinators in member states can request the temporary suspension of the service or online interface for an intermediary service.	Regulator has wide-ranging enforcement powers around failure to meet duties and requirements (confirmation decisions), including the ability to: • compel corrective action or • issue fines of up to the greater of: o 10 per cent of annual global turnover or o £18 million. In the most extreme cases, with the agreement of the courts, the Regulator will be able to require payment providers, advertisers, and internet service providers to stop working with a Service, preventing it from generating money or being accessed from the UK.	[Proposed penalties] **Administrative pecuniary penalties for non-compliance** The greater of: • 6 per cent of the gross global revenue of the person that is believed to have committed the violation or • $10 million A separate violation will apply for each day on which the violation continues. **Penalties on offence conviction** For operators, the greater of: • 8 per cent of the gross global revenue or • $25 million. For other persons: • Individuals up to $50,000 • Non-individuals, the greater of: o 3 per cent of the gross global revenue or o $10 million.

Table 4.2 Comparative penalties under Australian laws

^ Penalty unit values are calculated as at 22 April 2024

	Australian Consumer Law	Section 474.34 of the Criminal Code -Removing, or ceasing to host, abhorrent violent material	Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023	*Privacy Act 1988*	*Anti-Money Laundering and Counter-Terrorism Financing Act 2006*
Maximum Fines / Penalties	Penalties for unconscionable conduct, making false or misleading representations and supplying goods/services that do not comply with safety standards (or which are banned): For corporations, not more than the greater of: • $50,000,000 or 3 times the value of the 'reasonably attributable benefit' of the offence, or • If the court can determine the value of the benefit that the body corporate obtained, 3 times the value of the 'reasonably attributable benefit' of the offence, or • If the court cannot determine the value of that benefit, 30 per cent of the 'adjusted turnover' of the body corporate during the 'breach turnover period'. For a person other than a body corporate: • $2,500,000.	Penalties for an offence under this provision For corporations, the greater of: • 50,000 penalty units ($15.65 million), or • 10 per cent of the annual turnover of the body corporate during the period of 12 months ending at the end of the month in which the conduct constituting the offence occurred. For individuals: • 3 years imprisonment, or • 10,000 penalty units ($3.13 million), or • Both.	[Proposed penalties] ***Non-compliance with a code*** For corporations, the greater of: • 10,000 penalty units ($3.13 million), or • 2 per cent of global turnover For individuals • 2,000 penalty units ($626,000). ***Non-compliance with a standard*** For corporations, the greater of: • 25,000 penalty units ($7.825 million), or • 5 per cent of global turnover For individuals: • 5,000 penalty units ($1.565 million).	Serious and repeated interferences with privacy (civil penalty). For a body corporate, not more than the greater of: • $50,000,000, or • If the court can determine the value of the benefit that the body corporate obtained, 3 times the value of the reasonably attributable benefit of the contravention, or • If the court cannot determine the value of that benefit, 30 per cent of the 'adjusted turnover' of the body corporate during the 'breach turnover period'. For a person other than a body corporate: • $2,500,000.	AUSTRAC Civil penalties for designated services. For corporations: • 100,000 penalty units ($31.3 million). For individuals: • 20,000 penalty units ($6.26 million).

Investigation and information gathering powers

Most regulators are provided with investigation powers and information gathering powers. These powers can require a regulated entity to provide access to data necessary to assess compliance. Investigators can also be given powers to summon a person to appear and give evidence or to provide access to information to allow

for an investigation. These powers may be in addition to transparency reporting and independent audit obligations for service providers.

In Australia, in addition to powers to require transparency reporting in relation to the Basic Online Safety Expectations, the Commissioner has powers to investigate complaints or suspected breaches of the codes or standards. The Commissioner can investigate in response to complaints made under the complaints or content-based removal schemes and may also initiate investigations in relation to complaints, or suspected breaches of codes or standards under the Online Content Scheme. Investigation powers include powers to summon a person to attend before the Commissioner to answer questions, to provide information or documents to the Commissioner, and to examine a person under oath or affirmation.

The Commissioner also has powers to obtain identity information or the contact details of an end-user of a social media service, relevant electronic service, or designated internet service where there are reasonable grounds to believe the information is relevant to the operation of the Act.[55] Based on eSafety's experience in exercising this power, the 'identity' information obtained is often of limited utility due to the scant data collected by many services. For example, some services collect only IP information and an email address, which can lead to an investigative 'dead end'. User information is of most assistance when it includes telephone numbers and/or financial information.

Part 4 – Penalties, and investigation and information gathering powers – consultation questions

17. Does the Act need stronger investigation, information gathering and enforcement powers?
18. Are Australia's penalties adequate and if not, what forms should they take?
19. What more could be done to enforce action against service providers who do not comply, especially those based overseas?
20. Should the Commissioner have powers to impose sanctions such as business disruption sanctions?

---

[55] *Online Safety Act 2021*, section 194.

Part 5 – International approaches to address online harms

Combatting online harm is a global issue. Many large digital product and service providers are based overseas and the markets that they operate in are regulated by multiple governments. Jurisdictions around the world continue to examine the issue of how to keep their citizens safe online. Aligning approaches and sharing best practice, where appropriate in the Australian context, supports global consistency in regulating digital spaces, a benefit recognised by G20 leaders in 2023.[56]

While Australia has been a world leader in online safety regulation, the global regulatory environment is rapidly evolving, with newer regulatory schemes focusing on systemic protections, rather than episode-based interventions for particular types of online content. A variety of regulatory approaches exist internationally, with countries tending to take either a content and individual complaint-based approach (like Fiji) or a broader systemic approach such as the EU and UK which focus on systems and processes. Some governments (including in Australia and the Republic of Korea), have taken a hybrid approach, with the ability for individuals to make complaints about specific types of online content, as well as placing systemic requirements on digital platforms. Canada has also proposed a hybrid model under its draft Online Harms Act, and Ireland has indicated it may move to a hybrid model in the future.

Not all countries regulating online safety have nationally legislated regulatory frameworks. Countries such as Japan and New Zealand have voluntary codes, with specific harmful or illegal online content such as child sexual exploitation material covered by existing criminal laws. While most countries take a national response to online safety, some countries such as the United States have taken a state-based approach in responding to online service issues including age assurance.

Global trends

Since the Act came into effect, several jurisdictions have introduced or proposed new online safety regulatory frameworks. The UK's Online Safety Act 2023 (UK OSA), the EU's Digital Services Act (DSA), Ireland's Online Safety and Media Regulation Act 2022, Singapore's Online Safety (Miscellaneous Amendments Act) 2022 and Canada's proposed Online Harms Act aim to make the internet safer by focusing on systems and processes. While these more recent regulatory approaches are still being implemented and considered, they take different regulatory approaches in terms of scope, specificity, and the obligations imposed on digital platforms. This paper sets out the regulatory approach being adopted in the EU, UK and Canada and other jurisdictions in Appendix 2.

Evolving concepts in online safety regulation

Statutory duty of care approach

A statutory duty of care approach places duties on the entities who control and are responsible for a hazardous environment to achieve a desired outcome (harm prevention). This places a regulatory burden on the entity controlling the regulated environment, and can increase a regulatory framework's capacity to adapt to unique features and changes in the environment.

---

[56] G20 New Delhi Leaders' Declaration New Delhi, India, 9-10 September 2023, New Delhi 2023 G20 Leaders' Declaration (mea.gov.in).

> **The pace of change in both technology and behaviour on social media is such that detailed rules tackling specific harm are likely to become outdated or ineffective very quickly. Requiring operators to identify hazards and risk of harm avoids this problem.**
>
> Carnegie UK submission to House Select Committee on Social Media and Online Safety[57]

A statutory duty of care includes an overarching obligation to exercise care in relation to user harm (including through risk assessments and implementing mitigation measures). There is also an obligation to continually assess the effectiveness of those measures. The duties are enforceable, and penalties may apply for failure to comply.

Duties imposed through the UK's Online Safety Act provide an example of a statutory duty of care approach. Details on measures that comply with the duties will be provided in codes (which may need to cover a broad spectrum of regulated service models), but service providers can elect to implement alternative measures that meet duty requirements. The duties do not create a private right of claim for end-users, unlike the proposed Kids Online Safety Act in the United States or a common law action for negligence where end-users can take direct action against platforms for breach of a duty of care.

In Australia, an example of a statutory duty of care is provided in model Work Health and Safety laws which place a primary obligation on persons conducting a business or undertaking to ensure the health and safety of workers and others who may be affected by the work, so far as is reasonably practicable. Duties are imposed on persons who influence the way work is carried out and the integrity of products used, and officers are required to exercise 'due diligence' to ensure compliance. The model laws also recognise that more than one person can concurrently have the same duty, requiring each duty-holder to discharge their duty to the extent of their influence and control over the matter.

Submissions to Australia's 2022 House of Representatives Select Committee on Social Media and Online Safety inquiry queried whether Australia's regulatory models adequately protect users from harm. A number of submissions favoured placing a legally enforceable duty of care on social media platforms and other digital services.[58] The Committee found that a statutory duty of care model has 'significant strengths, and flips the onus of responsibility to provide and ensure user safety back onto social media platforms.'[59] The Committee also acknowledged the contribution of the Commissioner's Safety by Design program to improving user safety. As an enhancement to the Basic Online Safety Expectations, the Committee supported introducing a formal statutory duty of care framework that incorporates penalties for non-compliance and is modelled on the best interests of the child principle.[60] The Committee also acknowledged that the Act was still new, recommending a duty of care model be considered as part of a review of the Act.[61]

Best interests of the child principle

The 'best interests of the child' principle is set out in the United Nations Convention on the Rights of the Child. Article 3.1 makes the best interests of the child (everyone under 18 years) a primary consideration in actions and decisions concerning children.

---

[57] Carnegie UK 2022 Submission to the House Select Committee on Social Media and Online Safety, January 2022, [12] available at Submissions – Parliament of Australia (aph.gov.au).

[58] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.24], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[59] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.82], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[60] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.83],[5.86], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[61] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.84]-[5.85], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

> **Article 3.1**
>
> **In all actions concerning children, whether undertaken by public or private social welfare institutions, courts of law, administrative authorities or legislative bodies, the best interests of the child shall be a primary consideration.**
>
> United Nations Convention on the Rights of the Child

Further information on the best interests of the child principle is available through the Australian Human Rights Commission.

Submissions to the House of Representatives Committee on Social Media and Online Safety suggested the best interests of the child principle as a focus for online safety reforms.[62] The National Children's Commissioner favoured a requirement for social media and other online services to demonstrate that their services meet the best interests of the child principle, including 'considerations of privacy, security of personal data, protection from harm, a voice to express their views and the ability to seek, receive and convey information.'[63]

Children increasingly rely on online services in their everyday lives. Despite the benefits these services provide, there are concerns about the impact on children from spending large amounts of time online including in relation to their physical and mental health and emotional and social wellbeing.[64] There are also concerns that children are increasingly being 'datafied,' with potentially millions of data points collected on their location, interests, activities and moods.[65] In February 2023, the Attorney-General's Department released the Privacy Act Review Report which raised the need to better protect children's privacy online. The report included proposed privacy protections that would strengthen privacy protections for children and people experiencing vulnerability. As part of the Government response to this report, the Government will also introduce a Children's Online Privacy code which would apply to online services that are likely to be accessed by children.[66] The requirements of the code could help clarify how the bests interest of the child should be upheld in the design of online services.

Globally, regulators have introduced specific measures protecting children's online safety and rights.

- The UK's Online Safety Act imposes duties on regulated services with a dual purpose of identifying, mitigating and managing risk of harm from illegal content and activity, and 'content and activity that is harmful to children.' Specific duties apply to assess the likelihood of children accessing the service, and for services likely to be accessed by children. These include duties to conduct children's risk assessment and to protect children's' online safety.
- The EU's Digital Services Act includes specific requirements for very large online platforms and search engines to consider the rights of the child in conducting risk assessments for their services and systems, to consider risk mitigation measures to protect the rights of the child (where appropriate), including age verification and parental control tools, and tools aimed at helping minors signal abuse or obtain support. Examples of risks that may arise include 'the design of online interfaces which

---

[62] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.31], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[63] Social Media and Online Safety, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' (March 2022), [5.32], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[64] Commonwealth of Australia (2021) Physical activity and exercise guidelines for all Australians, accessed 26 April 2024.

[65] Australian Government, Attorney-General's Department, *Privacy Act Review Report 2022*, Privacy Act Review Report 2022 (ag.gov.au), 46, accessed 26 April 2024.

[66] Australian Government (2023), *Government Response – Privacy Act Review Report*, Government Response - Privacy Act Review (ag.gov.au), 13, accessed 26 April 2024.

intentionally or unintentionally exploit the weaknesses and inexperience of minors or which may cause addictive behaviour.'

- Canada's proposed Online Harms Act would impose a duty on regulated social media services to protect children in respect of the services that they operate by integrating specified features into the service design. The features would be specified in regulations.
- In Ireland, the regulator can issue guidance materials relating to harmful online content, matters addressed in their Act or in an online safety code, and 'otherwise for the protection of minors and the general public from harmful online content and age-inappropriate content.'

Australia's Act requires the Commissioner, where appropriate, to have regard to the Convention on the Rights of the Child in performing functions conferred by or under the Act, and in relation to children resident in Australia.[67] A new expectation has also been proposed as part of the Basic Online Safety Expectations for regulated service providers to ensure that 'the best interests of the child are a primary consideration in the design and operation of services likely to be accessed by children.' Public consultation on proposed amendments to the Basic Online Safety Expectations closed in February 2024. If implemented, service providers may be required to provide reports to the Commissioner in relation to the best interests of the child expectation, but the expectation would not create a legally enforceable duty.

Safety by Design

Safety by Design focuses on the ways that technology companies can minimise online threats by anticipating, detecting and eliminating online harms before they occur. There are three underlying principles:

- **Service provider responsibility** (the burden of safety should never fall solely on the service user)
- **User empowerment and autonomy** (products and services should align with the best interests of users)
- **Transparency and accountability** (ensuring platforms and services are operating according to their published safety objectives, sharing safety innovations and educating and empowering users).

In Australia, eSafety published a Safety by Design overview in 2019.[68] It is also a core expectation under the Basic Online Safety Expectations 'that the provider of the service will take reasonable steps to ensure that end-users are able to use the service in a safe manner.'[69] However, while related transparency reporting is enforceable, the Basic Online Safety Expectations do not create a legally enforceable duty. Safety by Design principles and tools, although voluntary, can also be used by industry participants as a way to support compliance with the Phase 1 Industry Codes.

Safety by Design is referenced in the UK's Online Safety Act, which states that the mandated duties imposed on service providers seek to secure Safety by Design.[70] The principle may also be considered for inclusion in a future iteration of Ireland's online safety codes with a possible requirement for platforms and services to undertake a Safety Impact Assessment.[71]

The underlying principles of Safety by Design are implied in duties outlined in Canada's proposed Online Harms Act, including the duty to protect children and to publish a digital safety plan. Similarly, the EU's Digital Services Act includes obligations for online platforms to ensure their interfaces meet certain design and

---

[67] *Online Safety Act 2021*, section 24.

[68] eSafety Commissioner, Safety by Design Overview, May 2019, SBD - Overview May19.pdf (esafety.gov.au), downloaded 28 February 2024. Published to promote online safety under the *Enhancing Online Safety Act 2015* (legislation replaced by the *Online Safety Act 2021*).

[69] *Online Safety Act 2021*, section 46(1)(a); Online Safety (Basic Online Safety Expectations) Determination 2022, section 6(1).

[70] *Online Safety Act 2023* (United Kingdom), section 1(3)(a).

[71] Coimisiún na Meán (2023) Call For Inputs: Online Safety, accessed 26 April 2024.

accessibility standards, enabling users to make free and informed decisions, and to implement measures to ensure a high level of security, privacy, and safety of minors.

Stronger enforcement powers

The UK's Online Safety Act provides the regulator with a spectrum of enforcement powers, including a provisional notice of contravention (that either provides steps required to remedy the non-compliance, or provides an opportunity to object to a proposed penalty), confirmations of contravention with associated remedial actions required, or penalties imposed. A daily rate penalty may apply where a regulated service fails to take required remedial action by the compliance date. Offences may apply for failing to comply with remedial actions required under a confirmation of contravention, a child's online safety duty, or a requirement related to child sexual exploitation or abuse. Offences carry a maximum penalty of two years imprisonment, or a fine, or both. The regulator can also seek a service restriction order from the courts if a service continually fails to comply with an enforceable action, or an access restriction order if the service restriction order was not sufficient to prevent significant harm arising from the failure. Senior managers of regulated services may also be liable for offences if the service fails to comply with information requirements.

Canada's proposed Online Harms Act enables the regulator to impose a penalty and to order corrective action if a person has committed a violation. Continued violations would be treated as a separate violation for each day the violation continues. Offences also apply to operators and other people who contravene the proposed Online Harms Act (financial penalty on conviction).

Transparency and data access

In Australia, the Basic Online Safety Expectations encourage the prevention of online harms by regulated service providers and have improved the transparency of actions taken. The Commissioner can require providers of a social media service, relevant electronic service, or designated internet service to report about their compliance with specific basic online safety expectations, providing greater transparency of online safety practices to the government and the community. While the expectations are voluntary, compliance with a reporting requirement, including in the manner and form specified, is not. The Commissioner has the power to publish summaries of the information received in the reports, and through this reporting has delivered greater transparency not otherwise achieved through voluntary transparency initiatives. As the Commissioner may prepare, and where appropriate publish, a statement of the service provider's compliance or non-compliance with the expectations, the current cost for non-compliance with the expectations is reputational. The Government response to the Privacy Act Review Report acknowledged the importance of transparency measures for providing greater transparency to individuals and for assisting regulators to ensure compliance with the relevant Act.[72]

Annual reporting provides transparency about eSafety's operations and regulatory actions. eSafety publishes its annual reports as a primary mechanism of accountability to the Australian Parliament. This includes reporting against 43 metrics, including informal removal requests (whether based on terms of service or legislation), formal removal notices (based on legislation), and reviews. eSafety also publishes regulatory guidance to help stakeholders understand eSafety's focus, approach and reasoning.

Jurisdictions overseas are also introducing transparency measures. These include regulatory powers to request information, mechanisms to provide authorised researchers with access to data, requiring publication of online safety risk assessments or risk mitigation measures, and mandatory audit requirements.

The EU's Digital Services Act provides a framework for compelling access to data from very large online platforms and very large online search engines for accredited researchers. Those very large services must also

---

[72] Australian Government (2023), *Government Response – Privacy Act Review Report*, Government Response - Privacy Act Review (ag.gov.au), accessed 26 April 2024.

provide regulators with data access on request to enable compliance monitoring, and at least once a year they must have an independent audit undertaken (which includes access to data and consideration of research reports) and report on content moderation activities.

Transparency measures in the UK focus on the highest risk or highest reach services, including for transparency reporting and requiring the publication of certain online safety information. The regulator can request information from regulated (and ancillary) services at any time, for the purpose of exercising or deciding to exercise online safety functions. Offences apply for failure to comply. The requested information could include, for example, service risk assessments and mitigation measures, or documents from engineers regarding new features. The highest reach and risk services will be required to provide a transparency report each year.

The UK also requires service providers to publish children's risk assessments and, depending on the service provided, may require the publication of risk assessments related to the impacts of safety measures and policies on user privacy and freedom of expression (Category 1 user-to-user services), the most recent illegal content risk assessment (Category 2A search services), or a summary statement of age verification and age assurance measures taken to ensure children are not normally able to access the service's content (services that provide pornographic content).

Proposed Online Harms legislation in Canada would establish a Commission with broad powers to summons, inspect, and hold hearings related to compliance with the legislation or certain content complaints. Service providers would be required to keep records of their compliance with the statutory duties, and publish an accessible digital safety plan addressing the duty to act responsibly. The Commission would also have powers to enable accredited persons to access data included in digital safety plan inventories for research, education, advocacy or awareness activities related to the purposes of the Act.

Supporting users

Australia's online safety regulatory framework provides two mechanisms in relation to internal processes for addressing certain online harms: industry codes and standards (in the case of illegal and restricted content) and the Basic Online Safety Expectations.

The industry codes include obligations on industry participants to receive and/or respond to user reports made to their service, including the code covering social media services. The Basic Online Safety Expectations Determination sets out non-binding expectations in relation to reports and complaints about a broader range of harms, including clear and identifiable mechanisms and policies and procedures to deal with reports and complaints about unlawful and harmful material. Some complaints about harmful online content and activity are beyond the remit of the Act, but may breach the service providers' terms of use (also known as 'terms of service'). In these circumstances, appropriate internal dispute resolution or complaint handling procedures can be lacking.

In September 2022, the fifth interim report of the Australian Competition and Consumer Commission (ACCC) Digital Platform Services Inquiry 2020–25 concluded that an 'ombuds scheme, and the ability to escalate complaints and disputes to an independent body is critical to ensuring the effectiveness of internal dispute resolution measures.'[73] The report recommended Australia establish an independent external dispute resolution scheme in the form of an ombuds scheme, as well as internal dispute resolution obligations.[74] The proposed ombuds would be able to compel information, make decisions that are binding on relevant

---

[73] Australian Competition and Consumer Commission (2022), *Digital platform services inquiry Interim report No. 5 – Regulatory reform* Digital platform services inquiry - September 2022 interim report.pdf (accc.gov.au), [4.3.1].

[74] Australian Competition and Consumer Commission (2022), *Digital platform services inquiry Interim report No. 5 – Regulatory reform* Digital platform services inquiry - September 2022 interim report.pdf (accc.gov.au), [4.3].

digital platforms, order compensation where appropriate, and investigate and refer systemic issues identified to regulators.[75] While the Inquiry is focused on competition and consumer issues, the report noted potential for an ombuds scheme to cover broader online disputes, including those related to privacy and online harms.[76]

The Government has committed to further work to develop internal and external dispute resolution requirements by calling on industry to develop voluntary internal dispute resolution standards by July 2024.

Having clear and transparent dispute resolution processes can improve a users' experience and provide transparency around content moderation decisions. Jurisdictions overseas have introduced a range of regulatory measures to address concerns about transparency of content moderation and dispute resolution processes. These include requiring online services to publish clear and accessible dispute resolution processes (UK), consider proportionate content moderation measures with risk mitigation measures (UK), to provide content moderation appeal mechanisms (Canada), or to publicly report on content moderation outcomes (EU). Canada is also proposing an Ombudsperson to provide independent guidance to online users, and the UK legislation allows for an alternative dispute resolution procedure to be imposed on Category 1 services.

Respecting human rights

In regulating the online environment, governments must consider how to uphold a range of fundamental human rights and supporting principles, including:

- the principle of the best interests of the child
- the principles of dignity, equality, and mutual respect
- the right to freedom of information, opinion, and expression
- the right to freedom of association
- the right to privacy
- the right to protection from exploitation, violence, and abuse
- the right to non-discrimination.[77]

There are important nuances to be considered in assessing human rights impacts. For example, legislative limits on permissible online activity can have the effect of restricting freedom of expression for some, while supporting safe freedom of expression for others who might otherwise be silenced by abuse or hate.

Similarly, where anonymity and identity shielding can protect the privacy and safety of online users, it can also be used to control and abuse people, and make it difficult to hold individuals to account.[78] Conversely, intentionally exposing an individual's identity, private information or personal details without their consent (doxxing) can undermine a person's privacy, security, safety, or reputation.

---

[75] Australian Competition and Consumer Commission (2022), *Digital platform services inquiry Interim report No. 5 – Regulatory reform* Digital platform services inquiry - September 2022 interim report.pdf (accc.gov.au), [4.3], accessed 26 April 2024.

[76] Australian Competition and Consumer Commission (2022), *Digital platform services inquiry Interim report No. 5 – Regulatory reform* Digital platform services inquiry - September 2022 interim report.pdf (accc.gov.au), [4.3.1], accessed 26 April 2024.

[77] Global Online Safety Regulators Network, *Position Statement: Human Rights & Online Safety Regulation* September 2023, 2.

[78] eSafety Commissioner, *Anonymity and identity shielding online - Tech trends position statement* January 2021**,** Anonymity and identity shielding | eSafety Commissioner, accessed 19 February 2024.

Part 5 – International approaches to address online harms – consultation questions

21. Should the Act incorporate any of the international approaches identified above? If so, what should this look like?
22. Should Australia place additional statutory duties on online services to make online services safer and minimise online harms?
23. Is the current level of transparency around decision-making by industry and the Commissioner appropriate? If not, what improvements are needed?
24. Should there be a mechanism in place to provide researchers and eSafety with access to data? Are there other things they should be allowed access to?
25. To what extent do industry's current dispute resolution processes support Australians to have a safe online experience? Is an alternative dispute resolution mechanism such as an Ombuds scheme required? If so, how should the roles of the Ombuds and Commissioner interact?
26. Are additional safeguards needed to ensure the Act upholds fundamental human rights and supporting principles?

Part 6 – Regulating the online environment, technology and environmental changes

Online harms can occur wholly or partially online and damage an individual's social, emotional, psychological, financial or physical safety.[79] Harms can occur in the production, distribution, or consumption of online content (content harms), as a result of online interactions with others (contact harms), or through behaviour facilitated by technology (conduct harms).[80] Existing harms are likely to be impacted and new challenges and harms are likely to emerge as new technologies become more readily available to consumers, and Australians change how they interact online, and the structure of online services evolve.

Regulatory opportunities and challenges continue to evolve at pace. Some emerging regulatory challenges include the creation of potentially harmful synthetic material using generative artificial intelligence, privacy challenges such as deliberate publication of people's personal or identifying information (doxxing) or regulating encrypted environments, issues arising from exposure to materials promoting self-harm, and new types of harms arising from interactions in immersive environments which may include augmented reality, virtual reality or mixed reality.

Regulatory frameworks are evolving to adapt to the scale and speed of harms arising online. While the global regulatory environment is trending toward harm mitigation through the largest or highest risk online service providers, decentralised online platforms and services (including Web 3.0 technologies) could introduce new regulatory challenges in online safety.

Online harms which may not be fully addressed under the Act

Online technology, the way Australians interact online, and online harms are constantly changing. Examples of emerging harms, including those arising from new and emerging technology are outlined below.
While regulatory frameworks cannot address every potential online harm, the review provides an opportunity to consider whether there are new or emerging harms that should be specifically addressed.

Cyber-flashing

Cyber-flashing refers to online actions involving a user sending or sharing nude, semi-nude or sexual photos or videos without the recipient's consent. Cyber-flashing as a topical and emerging harm was explored through the 2023 Online Safety Issues Survey.

The survey found that almost 8 per cent of those surveyed had experienced cyber-flashing over a 12-month period. Those aged 35-54 (10 per cent) and those cohorts who speak a language other than English at home, live with disability, identify as LGBTQIA+, or identify as Aboriginal and or Torres Strait Islander were generally more likely to have been cyber-flashed.[81]

Both the UK and Ireland created offences related to cyber-flashing in their online safety regulatory frameworks. The UK's Online Safety Act created a criminal offence for a person sending photographs or films of their genitals to cause distress. In Ireland, a new criminal offence applies in relation to online content where a person exposes their genitals intending to cause fear, distress or alarm to another person.

---

[79] eSafety Commissioner (2022), *Australia's eSafety Strategy 2022-25*, eSafety Strategy 2022-25.pdf, 5, accessed 26 April 2024.
[80] World Economic Forum, Typology of Online Harms, 2023, 5, WEF_Typology_of_Online_Harms_2023.pdf. accessed 26 April 2024.
[81] Social Research Centre (2023), *2023 Online Safety Issues Survey – Summary report*, 2023 Online Safety Issues Survey, accessed 26 April 2024.

Online hate

There are different views about what constitutes hate speech, which can be seen as falling anywhere on a spectrum of harm from offensive and insulting on the lower end to seriously menacing, threatening and harassing on the higher end. Hate speech is not new, but its prevalence can be easily spread online at a magnitude and order not seen before, potentially having a greater detrimental impact on social cohesion.

> **hate speech…can be disseminated like never before, worldwide, in a matter of seconds, and sometimes remain persistently available online**
>
> *Delfi AS v Estonia* App. no. 64569/09 ECHR, 16 June 2015, [110]

The Act does not define online hate or confer specific hate speech-related powers on the Commissioner but provides some protections through its regulatory schemes. Some individual communications may fall under the adult cyber-abuse, child cyberbullying, or image-based abuse complaints schemes. The Online Content Scheme for Class 1 material includes online content that counsels, promotes, encourages, urges, instructs or praises the doing of a terrorist act. It also includes material that promotes, incites or instructs in matters of crime. The Basic Online Safety Expectations also encourage industry to ensure services are safe for Australians and require greater transparency around services' safety measures, including measures to enforce their terms of use which usually prohibit the posting of hate speech.

All Australian jurisdictions have frameworks to deal with hate speech (online and offline), through anti-discrimination, anti-vilification and incitement laws. The *Racial Discrimination Act 1975* is the only federal anti-discrimination law with a hate speech provision. Section 18C makes it unlawful to do an act, otherwise than in private, which is 'reasonably likely' to offend, insult, humiliate, or intimidate another person or group on the basis of their race, colour or national or ethnic origin. This 'racial hatred 'is treated as a civil wrong.

Although hate speech is not specifically criminalised under federal law, other criminal laws may apply. These include offences for urging violence, using a postal or carriage service to menace, harass or cause offence, and advocating terrorism.

In February 2024, the Standing Council of Attorney's-General acknowledged an increased prevalence of vilification, particularly online across social media platforms. The Council also noted the Government's intent to progress legislative reforms to strengthen protections against vilification and hate speech.[82]

Australia's Online Safety Act could be amended in a variety of ways to complement broader Government measures addressing online hate. Options could include, for example, expanded complaints-based schemes or further obligations or expectations in relation to online services' systems or processes. Globally, measures to address concerns over inappropriate content moderation have included enhanced content moderation transparency (process or reporting), mandatory appeal mechanisms for the people who posted moderated content, or using external dispute resolution frameworks.

During the consultation for the National Anti-Racism Framework, the Australian Human Rights Commission heard that while online platforms can foster positive and inclusive spaces, they are often spaces where racism and dehumanisation occur and misinformation is spread.[83] Some online platforms have policies around hate speech and users can report it to the service. However, hate speech is highly contested and context dependent, and these policies are not always enforced in line with community expectations. The Australian Human Rights Commission has also acknowledged the lengthy and challenging process for seeking redress,

---

[82] Standing Council of Attorney's-General Communiqué, 23 February 2024, Standing Council of Attorneys-General - 23 February 2024 (ag.gov.au).

[83] Australian Human Rights Commission (2022), *National Anti-Racism Framework Scoping Report*, National Anti-Racism Framework Scoping Report 2022, 131, accessed 26 April 2024.

'especially for self-represented complainants, given the lack of explicit coverage for religious identities, the 6-month limitation period, and difficulties and costs associated with progressing complaints to the Federal Court and Federal Circuit and Family Court of Australia if the conciliation process is unsuccessful.'[84]

Volumetric (pile-on) attacks

Volumetric attacks (or 'pile-on attacks'), occur where a person is tagged or linked to an abusive post which others like, share, or repost with additional commentary. Often the content is shared with an accelerating level of outrage and toxicity, and ultimately a high volume of abuse. Volumetric attacks often involve abusive posts connected with the target, which others like, share, or repost with additional commentary, and they sometimes involve coordinated and/or disingenuous behaviour. Volumetric attacks can be among the most serious forms of online abuse.[85] The harmfulness of individual instances of conduct can be damaging to the targeted user's wellbeing, and when done on an extensive scale through volumetric attacks, has the ability to magnify and compound the adverse impacts on end-users. The distribution of harmful conduct between individual users and content across platforms can also mean there is no single point for regulatory action.

The Act currently requires each individual post to be assessed against the threshold for regulatory action under the child cyberbullying or adult cyber-abuse schemes. However, while individual posts considered in isolation may not meet the regulatory thresholds, the volume and speed of pile-on attacks can amplify the harm. The intention to cause a volumetric attack, or the fact that a volumetric attack has occurred, may be a relevant consideration when investigators are assessing whether an individual post that has been reported meets the threshold for regulatory action.

The Basic Online Safety Expectations also include an expectation that services consult and cooperate with other services to promote user safety, including a reasonable step of detecting volumetric or cross-platform attacks. However, platform design, including recommender systems, can influence the nature of online communications by favouring incendiary or extreme content.[86] Social media platforms can also amplify expressions of moral outrage over time through users receiving increased numbers of 'likes' and 'shares' for their content.[87]

Technology-facilitated abuse

'Technology-facilitated abuse' is 'using technology to enable, assist or amplify abuse or coercive control of a person or group of people.'[88] It can include any form of abuse that is enabled through digital technologies, including online. This includes where technology is used as part of stalking or monitoring, psychological and emotional abuse (including threats), sexual violence or harassment, bullying or hate speech. Specific forms of technology-facilitated abuse include cyber abuse and image-based abuse.

---

[84] Australian Human Rights Commission, *National Anti-Racism Framework Scoping Report*, National Anti-Racism Framework Scoping Report 2022, 151, accessed 26 April 2024.

[85] Parliament of the Commonwealth of Australia, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' March 2022, [2.24], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[86] Luke Munn (2020), Digital Cultures Institute, New Zealand, '*Angry by design: toxic communication and technical architectures*' https://doi.org/10.1057/s41599-020-00550-7. Angry by design: toxic communication and technical architectures | Humanities and Social Sciences Communications (nature.com).

[87] Bill Hathaway (2021), *'"Likes" and "shares" teach people to express more outrage online'*, YaleNews online, 'Likes' and 'shares' teach people to express more outrage online | YaleNews.

[88] World Economic Forum, *Typology of Online Harms*, 2023, 8 WEF_Typology_of_Online_Harms_2023.pdf (weforum.org).

A 2022 survey of Australian adults found that one in two Australians had experienced technology-facilitated abuse behaviours in their lifetime.[89] The likelihood was higher for LGB+ Australians, Aboriginal and Torres Strait Islander people, Australians aged 18-44 years, and Australians with a disability.[90] The survey also found that one in four Australians had perpetrated technology-facilitated abuse in their lifetime, with a higher likelihood of perpetration identified for LGB+ Australians, Aboriginal and Torres Strait Islander people, and Australians aged 18-44 years.

Technology-facilitated abuse is typically gendered in nature, particularly in the context of family, domestic and sexual violence. Australian women are significantly more likely to experience abuse perpetrated by a man than by another woman in their most recent experience. Women are also found more likely to:

- experience technology-facilitated abuse from an intimate partner or former partner
- report emotional and psychological impacts
- experience co-occurring abuse from the same perpetrator.[91]

The potential for digital technologies to cause harm is also highlighted in the *National Plan to End Violence against Women and Children 2022-2032*.[92]

'Technology-facilitated gender-based violence' is a subset of technology-facilitated abuse. It can include a spectrum of behaviours, such as stalking, bullying, harassment, including sexual harassment, defamation, hate speech, doxxing, image-based abuse or exploitation, where the abuse is based in gender. The behaviours can result in physical, sexual, psychological, social, political or economic harms or other infringements of rights and freedoms on the basis of gender characteristics.[93] This type of violence when targeted at women manifests differently to violence and abuse targeted at men. It tends to be violent, sexualised, and include threats of sexual violence and rape towards the woman and her children. It will often target a women's physical appearance, fertility and virtue. Cyberstalking is particularly insidious. Information about using spyware is readily available online. Apps available in Australia enable the user to track another person's location and track activities such as texts, calls and internet browsing, and may be undetectable on the owner's device.[94]

Regulatory schemes in the Act specifically address harms to individuals through adult cyber-abuse, child cyberbullying, and image-based abuse. The Commissioner has also produced a range of educational resources on technology-facilitated abuse and how to stay safer online.

More information on these resources can be found here.

---

[89] Australian National Research Organisation for Women's Safety (ANROWS), *Technology-facilitated abuse: National survey of Australian adults' experiences*, July 2022, 8, 4AP.3-Flynn-TFa3-Survey-of-VS.pdf (anrowsdev.wpenginepowered.com), accessed 26 April 2024.

[90] Australian National Research Organisation for Women's Safety (ANROWS), *Technology-facilitated abuse: National survey of Australian adults' experiences*, July 2022, 8-9, 4AP.3-Flynn-TFa3-Survey-of-VS.pdf (anrowsdev.wpenginepowered.com). LGB+ is used when reporting on research focussed on sexuality separately from gender-diverse populations.

[91] Australian National Research Organisation for Women's Safety (ANROWS), *Technology-facilitated abuse: National survey of Australian adults' experiences*, July 2022, 9, 4AP.3-Flynn-TFa3-Survey-of-VS.pdf (anrowsdev.wpenginepowered.com).

[92] The National Plan to End Violence against Women and Children 2022-2032 | Department of Social Services, Australian Government (dss.gov.au).

[93] World Economic Forum, Typology of Online Harms, 2023, 8 WEF_Typology_of_Online_Harms_2023.pdf (weforum.org).

[94] Anne Summers, *How tech became the next frontier in domestic violence*, The Saturday Paper, 16 March 2024, How tech became the next frontier in domestic violence | The Saturday Paper.

Online abuse of public figures

Public figures and those with a public profile are subject to high rates of online abuse and harassment and are often at greater risk of online abuse than everyday private individuals.[95] Among them, women and minority public figures are the most targeted, as well as civil society advocates and activists.[96]

Online abuse, which can include trolling, stalking, impersonation accounts, image-based abuse and sexual harassment, can have serious professional and personal impacts.[97] In several cases, online abuse of public figures has preceded suicide.[98] Online abuse may also force public figures to withdraw from public life, and stifle the quality of public debate by making it more difficult for public figures to participate safely in online discourse.[99] In the context of women journalists, this phenomenon has been coined 'the chilling effect', where the 'chilling' of women's active participation in public debate is described as a threat to the public's right to information and an attack on media freedom and democracy.

Public figures, such as journalists, sports people, or politicians often have a professional requirement to be active online and engage with a range of social media platforms. Given this dependence, they may not have the option to remove themselves from abusive online environments. High-profile exposure combined with the potential attention on the content they post, increases a public figure's risk of exposure to online abuse. High amounts of online abuse, that may compound into volumetric attacks but do not individually meet thresholds for adult cyber-abuse are not covered under the Act's current schemes, leaving targeted figures reliant on assistance from online services.

Platform policies are unclear about how they define public figures, and definitions across platforms are inconsistent. Where defined, platforms often provide fewer protections to public figures on the basis of freedom of expression or public interest. Most policies of larger platforms reflect a higher threshold for addressing online harms directed at public figures than everyday users. Often, platforms do not differentiate between different types of public figures and fail to acknowledge the varying levels of resources and support available to different types of public figures.[100]

Body image harms / Self-harm promotion

Online content has the capacity to encourage destructive or unhealthy behaviours for online users.[101] A 2022 study from eSafety showed that almost two in five (37 per cent) of young people aged 14 to 17 were exposed to potentially harmful online content related to drug taking in the past year and 28 per cent had been exposed to content promoting unhealthy eating. In addition, 25 per cent of young people aged 14 to 17 were exposed to self-harm content and 20 per cent were exposed to content about ways to take their own life. The

---

[95] Rob Cover, Henry N, Gleave J, Greenfield S, Grechyn V (2024), '*Protecting Public Figures Online: How Do Platforms and Regulators Define Public Figures?*', Media International Australia, 0(0):1-15.

[96] Ghaffari S (2022), '*Discourses of celebrities on Instagram: digital femininity, self-representation and hate speech'*, Critical Discourse Studies, 19(2):161-178.

[97] eSafety Commissioner (2023), '*What is online abuse?*' accessed 19 February 2024.

[98] Rob Cover et al (2024), '*Protecting Public Figures Online: How do Platforms and Regulators Define Public Figures*?', Media International Australia, 2.

[99] Rob Cover et al (2024), '*Protecting Public Figures Online: How do Platforms and Regulators Define Public Figures?*', Media International Australia, 3.

[100] In circumstances where public figures are supported by employers or others, it is then the supporting individuals who are experiencing the harmful content in place of, or in addition to the public figure.

[101] Parliament of the Commonwealth of Australia, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' March 2022, [2.41], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

research found that girls were more likely to have been exposed to discussions related to taking drugs, unhealthy eating and ways to be thin, or ways to physically harm or hurt themselves.[102]

Submissions to the House of Representatives Select Committee on Social Media and Online Safety raised concerns about the capacity for online content to encourage or promote destructive or unhealthy behaviour for users, with examples including self-harm, suicidal ideation and content promoting or instructing in disordered eating behaviour.[103]

Algorithms have been criticised for escalating exposure to harmful content. A 2022 study from The Center for Countering Digital Hate found that new teen accounts on TikTok were recommended eating disorder and self-harm content within minutes of scrolling the App's 'For You' feed.[104] Research from the UK in 2024 identified personalised social media feeds as the most mentioned pathway to encountering harmful content. Participants reported that 'initial encounters were often unintentional, with children being algorithmically recommended content they had not sought out.'[105]

> **If a child is particularly vulnerable in the real world, being served up more and more content relating to self-harm and suicide, dangerous challenges, or body image and eating disorders could not only have negative mental health impacts, but also potentially place them in real physical danger.** [106]
>
> eSafety Commissioner - 7 December 2022

In January 2024, the UK's online safety regulator published research revealing how major search engines act 'as gateways to harmful self-injury related web pages, images and videos.'[107] One in every five self-injury search results linked to content celebrating, glorifying or instructing about harmful self-injury, suicide or eating disorders. The report also acknowledged the challenges that deliberately obscured search terms present in identifying harmful content.

Internationally, several countries have introduced provisions to specifically address the risk of exposure to content promoting suicide, self-injury or disordered eating.

- The UK's Online Safety Act sets out duties to protect children's online safety for services likely to be accessed by children, including risk assessments and processes and procedures that minimise or prevent children from accessing content that is harmful to children. Stronger obligations apply to 'priority content that is harmful to children', including content that encourages, promotes or provides instruction for suicide, an act of deliberate self-injury, or an eating disorder or associated behaviours. The content could be text, graphics, an emoji, or a symbol. Category 1 user-to-user platforms also have a duty to provide tools that empower adult users to minimise their exposure to content that encourages, promotes or provides instructions for suicide, deliberate self-injury, an eating disorder, or behaviours associated with an eating disorder.

---

[102] eSafety Commissioner (2022), *Mind the Gap - Parental awareness of children's exposure to risks online* Mind the Gap - Parental awareness of children's exposure to risks online - FINAL.pdf (esafety.gov.au), 47, accessed 26 April 2024.
[103] Parliament of the Commonwealth of Australia, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' March 2022, [2.41]-[2.43], Social Media and Online Safety – Parliament of Australia (aph.gov.au).
[104] Center for Countering Digital Hate (2022), *Deadly by Design – TikTok pushes harmful content promoting eating disorders and self-harm into users' feeds,* December 2022, CCDH-Deadly-by-Design_120922.pdf (counterhate.com), 7, accessed 26 April 2024.
[105] Ipsos UK and TONIC Research (2024), *Online Content: Experiences of children encountering online content relating to eating disorders, self-harm and suicide*, March 2024, Online content: Experiences of children encountering online content relating to eating disorders, self-harm and suicide (ofcom.org.uk), 31, accessed 26 April 2024.
[106] eSafety Commissioner (2022), *Time to take a look under the virtual hood at how algorithms might be harming our kids*, 7 December 2022, Time to take a look under the virtual hood at how algorithms might be harming our kids | eSafety Commissioner, accessed 26 April 2024.
[107] Ofcom (2024), Search engines can act as one-click gateways to self-harm and suicide content, Search engines can act as one-click gateways to self-harm and suicide content - Ofcom, accessed 26 April 2024.

- Ireland's safety codes will impose obligations on industry to address online content which, subject to a risk test, may be considered harmful, including content that promotes suicide, self-harm or eating disorders. To meet the relevant risk test the content must either give rise to a risk to a person's life or to a risk of significant harm to a person's physical or mental health, where the harm is reasonably foreseeable.
- Canada's proposed Online Harms Act would impose a duty on operators of social media services to implement measures to mitigate the risk of users being exposed to harmful content. Harmful content is defined to include 'content that induces a child to harm themselves', including by advocating self-harm, disordered eating or dying by suicide.

Proposed amendments to the Basic Online Safety Expectations Determination include a new expectation that recommender systems are designed and implemented in a manner that enables their safe use, and that services minimise the extent to which recommender systems amplify unlawful or harmful material. In addition, a new expectation has been proposed that the best interests of the child are a primary consideration in the design and operation of services likely to be accessed by children. However, the Basic Online Safety Expectations do not create a legally enforceable duty.

Beyond recommender systems, a 2024 report from Reset Australia concluded that all systems should be considered, finding that content recommender systems, content moderation systems, advertisement approval systems, and advertisement management systems could all create risk of exposure to pro-eating disorder content.[108]

Potential online harms and emerging technologies

The Act is technology-neutral, focusing primarily on specific online harms rather than the means through which the harm was generated. Evolving technologies can provide new functionalities and improve user experience but can also increase the frequency and speed with which harms can occur.

Generative artificial intelligence

'Generative artificial intelligence' describes the process of using machine learning to create digital content such as new text, images, audio, video and multimodal experience simulations. Examples include:

- Text-based chatbots or programs designed to simulate conversations with humans, such as Anthropic's Claude, Bing Chat, ChatGPT, Google Gemini, and Snapchat's My AI
- Image or video generators, such as the Bing Image Creator from Microsoft Designer, DALL-E 2, Midjourney, and Stable Diffusion
- Voice generators, such as Microsoft VALL-E.

The rapid deployment of generative artificial intelligence and the scale and sophistication of content produced, have the potential to amplify online harms. This could be realised through algorithmic bias resulting from automated decision-making or exposure to discrimination and bias through the outputs of generative artificial intelligence tools. Online harm examples include chatbots providing inappropriate and harmful responses to user prompts, the spread of hyper realistic generative artificial intelligence deepfakes, and the creation of synthetic child sexual abuse material.

---

[108] Reset. Australia (2024), *Not Just Algorithms: Assuring User Safety Online with Systemic Regulatory Frameworks*, Report: Not Just Algorithms – Reset Australia, accessed 26 April 2024.

Immersive technologies

Immersive technologies allow users to engage in a virtual world where users interact with each other in an immersive and interactive computer-generated environment. According to eSafety research, an estimated 680,000 adults in Australia may be engaging in the metaverse, with half of those interacting in these environments at least once a month. More than half of those engaging in the metaverse are using haptic technologies. Haptic technologies transmit tactile information through sensations such as vibration, touch and force feedback to enhance the user experience.

Virtual world interactions, especially when enhanced by haptic technologies, introduce the potential for new online harms that were previously limited to the physical world. In December 2023, the Digital Regulation Co-operation Forum (UK) identified potential issues arising from immersive technologies, including novel forms of harm, and the convergence of immersive social media, retail and gaming hindering redress.[109] The potential for novel harms was demonstrated in January 2024 when UK police were reported to be investigating a case of a child whose avatar was sexually assaulted in an immersive video game.

Recommender systems and algorithms

Recommender systems prioritise content or make personalised content suggestions to online service users. Recommender systems and their underlying algorithms are built into many online services, sorting through vast amounts of data to present content that is relevant to users.

For example, social media services use recommender algorithms to personalise what is suggested or promoted to users and to increase the reach of prioritised content and accounts. Online services use recommender systems to drive engagement and maintain 'stickiness', in order to encourage its users to spend more time on the service. However, this can also create an incentive to promote content that may be harmful but attention-grabbing, amplifying mis/disinformation, extremist views, and reinforcing perspectives that the user may already be aligned with, creating echo-chambers.

End-to-end encryption

End-to-end encryption is a means of securing communications from one end point to another and is an important defence against security breaches that would otherwise have serious consequences for online users. It transforms standard text, image, audio and video files, and live video streams, into an unreadable format while still on the sender's system or device. The content can only be decrypted and read once it reaches its final destination.

End-to-end encryption is increasingly being adopted by services which offer messaging functions to consumers. However, it can also conceal harmful conduct or hinder investigation of the distribution of harmful and illegal online content such as child sexual exploitation material.[110]

In July 2022, the Australian Institute of Criminology explored the potential impact of end-to-end encryption on the detection of child sexual abuse material.[111] The Institute's report noted the challenges end-to-end

---

[109] Digital Regulation Co-operation Forum (DRCF) (2023), *Immersive Technologies Foresight paper*, Immersive Technologies Foresight Paper | DRCF, accessed 26 April 2024.

[110] eSafety Commissioner (2023), *End-to-end encryption: Position statement,* End-to-end encryption trends and challenges — position statement | eSafety Commissioner, accessed 19 February 2024.

[111] Australian Institute of Criminology (2022), *Trends and issues in crimes and criminal justice* No. 653 July 2022 'Child sexual abuse material and end-to-end encryption on social media platforms: An overview', Child sexual abuse material and end-to-end encryption on social media platforms: An overview (aic.gov.au), accessed 26 April 2024.

encryption created for law enforcement investigations, and limitations placed on companies' ability to prevent, detect and report child sexual abuse material occurring on their platforms.[112]

Company variations in detection methods, evaluation of detection methods, and transparency reporting were identified as areas for improvement. The report also noted that statistical drops in rate of reporting over time coincided with service transitions to end-to-end encryption, concluding that adoption of end-to-end encryption by more electronic service providers would likely provide a haven for child sexual exploitation and abuse material offending, rather than preventing it.

In December 2023, Meta announced that end-to-end encryption would be enabled as a default on Messenger (previously opt-in)[113] generating concerns over the potential impact on detection of child sexual abuse material.[114] Apple already having end-to-end encryption on some of its services committed to implementing its Child Sexual Abuse Material (CSAM) detection in iCloud Photos, a privacy-preserving photo-scanning tool, but decided to not proceed. In 2022 Meta made around 27 million reports to the National Center for Missing and Exploited Children compared to 234 for Apple.[115]

In 2022, the House of Representatives Select Committee on Social Media and Online Safety concluded that while privacy concerns are critical to the rights of all internet users, those issues did not 'outweigh the fundamental issue of ensuring safety in online environments.'[116]

Changes to technology models (decentralised platforms)

There is growing interest in developing decentralised online platforms and services (sometimes referred to as Web 3.0 or DWeb). Decentralisation has the potential to provide users with more power online by reducing reliance on mainstream, centralised servers and distributing responsibility for data sharing and storage to communities of users. While this could provide greater control and information protection to users, it could also create online safety regulatory challenges. Within the current regulatory framework, decentralisation makes it more difficult to hold users responsible for illegal or harmful content and conduct.[117]

Existing decentralised services include peer-to-peer services, blockchain-based services, and federated services that run on independent servers, such as Mastodon. Decentralised services are typically created for the purpose of being censorship resistant. However, this raises concerns around the ability to moderate or regulate decentralised services or platforms, potentially increasing the vulnerability of marginalised individuals and groups or creating space for criminal activities or users who have been removed from mainstream services.

---

[112] Australian Institute of Criminology (2022), *Trends and issues in crimes and criminal justice* No. 653 July 2022 'Child sexual abuse material and end-to-end encryption on social media platforms: An overview' Child sexual abuse material and end-to-end encryption on social media platforms: An overview (aic.gov.au), accessed 26 April 2024.

[113] Josh Taylor (2023), *Meta begins rolling out end-to-end encryption across Messenger and Facebook,* The Guardian, 7 December 2023, Meta begins rolling out end-to-end encryption across Messenger and Facebook | Facebook | The Guardian, accessed 26 April 2024.

[114] Josh Taylor (2023), *TechScape: Will Meta's encryption plans be a 'devastating blow' to child safety online?,* The Guardian, 12 December 2023, TechScape: Will Meta's encryption plans be a 'devastating blow' to child safety online? | Meta | The Guardian, accessed 26 April 2024.

[115] Josh Taylor (2023), *TechScape: Will Meta's encryption plans be a 'devastating blow' to child safety online*? The Guardian, 12 December 2023, TechScape: Will Meta's encryption plans be a 'devastating blow' to child safety online? | Meta | The Guardian, accessed 26 April 2024.

[116] Parliament of the Commonwealth of Australia, House of Representatives Select Committee on Social Media and Online Safety, 'Social Media and Online Safety' March 2022, [5.75], Social Media and Online Safety – Parliament of Australia (aph.gov.au).

[117] eSafety Commissioner (2022), Decentralisation – position statement, Decentralisation – position statement | eSafety Commissioner, accessed 26 April 2024.

Regulatory governance models

In Australia, the Online Safety Act establishes the eSafety Commissioner as an independent statutory office holder with staff and support provided by the Australian Communications and Media Authority. In the UK, Ofcom administers the Online Safety Act and has a Board to provide strategic direction and serve as the main decision-making body while the Ofcom Executive run the organisation. Canada's proposed Online Harms Act would see a Digital Safety Commission, comprising of three to five full time members, established to administer the Act. Canada's proposed Online Harms Act would also see a Digital Safety Ombudsperson to support users and advocate for the public interest in dealing with online safety systemic issues.

While eSafety is government funded, other jurisdictions such as the UK and the EU (and proposed by Canada), require online services in certain circumstances to pay towards the cost of regulation ('cost recovery'). Further detail can be found in Appendix 2.

Part 6 – Regulating the online environment, technology and environmental changes – consultation questions

27. Should the Commissioner have powers to act against content targeting groups as well as individuals? What type of content would be regulated and how would this interact with the adult cyber-abuse and cyberbullying schemes?
28. What considerations are important in balancing innovation, privacy, security, and safety?
29. Should the Act address risks raised by specific technologies or remain technology neutral? How would the introduction of a statutory duty of care or Safety by Design obligations change your response?
30. To what extent is the Act achieving its object of improving and promoting online safety for Australians?
31. What features of the Act are working well, or should be expanded?
32. Does Australia have the appropriate governance structures in place to administer Australia's online safety laws?
33. Should Australia consider introducing a cost recovery mechanism on online service providers for regulating online safety functions? If so, what could this look like?

Part 7 - Summary of consultation questions included in this paper

Part 2 – Australia's regulatory approach to online services, systems and processes

1. Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?
2. Does the Act capture and define the right sections of the online industry?
3. Does the Act regulate things (such as tools or services) that do not need to be regulated, or fail to regulate things that should be regulated?
4. Should the Act have strengthened and enforceable Basic Online Safety Expectations?
5. Should the Act provide greater flexibility around industry codes, including who can draft codes and the harms that can be addressed? How can the codes drafting process be improved?
6. To what extent should online safety be managed through a service providers' terms of use?
7. Should regulatory obligations depend on a service providers' risk or reach?

Part 3 – Protecting those who have experienced or encountered online harms

8. Are the thresholds that are set for each complaints scheme appropriate?
9. Are the complaints schemes accessible, easy to understand and effective for complainants?
10. Does more need to be done to make sure vulnerable Australians at the highest risk of abuse have access to corrective action through the Act?
11. Does the Commissioner have the right powers to address access to violent pornography?
12. What role should the Act play in helping to restrict children's access to age inappropriate content (including through the application of age assurance)?
13. Does the Commissioner have sufficient powers to address social media posts that boast about crimes or is something more needed?
14. Should the Act empower 'bystanders', or members of the general public who may not be directly affected by illegal or seriously harmful material, to report this material to the Commissioner?
15. Does the Commissioner have sufficient powers to address harmful material that depicts abhorrent violent conduct? Other than blocking access, what measures could eSafety take to reduce access to this material?
16. What more could be done to promote the safety of Australians online, including through research, educational resources and awareness raising?

Part 4 – Penalties, and investigation and information gathering powers

17. Does the Act need stronger investigation, information gathering and enforcement powers?
18. Are Australia's penalties adequate and if not, what forms should they take?
19. What more could be done to enforce action against service providers who do not comply, especially those based overseas?
20. Should the Commissioner have powers to impose sanctions such as business disruption sanctions?

Part 5 – International approaches to address online harms

21. Should the Act incorporate any of the international approaches identified above? If so, what should this look like?
22. Should Australia place additional statutory duties on online services to make online services safer and minimise online harms?
23. Is the current level of transparency around decision-making by industry and the Commissioner appropriate? If not, what improvements are needed?
24. Should there be a mechanism in place to provide researchers and eSafety with access to data? Are there other things they should be allowed access to?

25. To what extent do industry's current dispute resolution processes support Australians to have a safe online experience? Is an alternative dispute resolution mechanism such as an Ombuds scheme required? If so, how should the roles of the Ombuds and Commissioner interact?
26. Are additional safeguards needed to ensure the Act upholds fundamental human rights and supporting principles?

Part 6 – Regulating the online environment, technology and environmental changes

27. Should the Commissioner have powers to act against content targeting groups as well as individuals? What type of content would be regulated and how would this interact with the adult cyber-abuse and cyberbullying schemes?
28. What considerations are important in balancing innovation, privacy, security, and safety?
29. Should the Act address risks raised by specific technologies or remain technology neutral? How would the introduction of a statutory duty of care or Safety by Design obligations change your response?
30. To what extent is the Act achieving its object of improving and promoting online safety for Australians?
31. What features of the Act are working well, or should be expanded?
32. Does Australia have the appropriate governance structures in place to administer Australia's online safety laws?
33. Should Australia consider introducing a cost recovery mechanism on online service providers for regulating online safety functions? If so, what could this look like?

Part 8 – Call for submissions and next steps

Call for submissions

The questions in this issues paper are provided as a guide only and are not intended to restrict your participation in the consultation process. Please respond to the questions of interest to you – you do not need to respond to every question. In providing your responses it will be helpful if you:

- where possible, identify which parts of the Act your comments relate to
- describe what is working well and what improvements can be made
- explain what the impact of these improvements would be on you, others and the online environment, and
- provide any available data, evidence or case studies to support your view.

Closing dates

Written submissions are due by **5pm (AEST) Friday 21 June 2024**.

How to make a submission

Written submissions can be lodged online at: https://www.infrastructure.gov.au/have-your-say/.

If you are unable to use the webform, please send your submission to OSAReview@COMMUNICATIONS.gov.au or post it to:

> Director – Strategy and Research
> Online Safety, Media and Platforms Division
> Department of Infrastructure, Transport, Regional Development, Communications and the Arts
> GPO Box 594 Canberra, ACT 2601

Questions about the submission process can be sent to: OSAReview@COMMUNICATIONS.gov.au.

Publication of submissions

All written submissions (other than private submissions) will be made publicly available by the department unless a respondent specifically requests that a submission, or part of a submission, be kept confidential. Comments will not be published, unless you clearly indicate you would like it to be published as a submission. Comments that are not published will still inform the review process in the same manner as submissions.

There are legal considerations relevant to what the department can publish on its website. Submissions that may expose the department to legal action will not be published. The department reserves the right not to publish any submission, or part of a submission, which in its view contains potentially offensive or defamatory material, or for confidentiality reasons.

The department is subject to the *Freedom of Information Act 1982* and comments and submissions may be required to be disclosed by the department in response to requests made under that Act.

What happens next?

The written submissions will inform the development of recommendations to the Minister in respect of proposed changes to the Act. The terms of reference require a report of the review to be provided to the Minister for Communications by 31 October 2024.

Appendix 1: Government actions against other online harms

The list below represents examples of other Australian Government protections against other online harms.

- **The Office of the Australian Information Commissioner** promotes and upholds privacy and information access rights, such as those provided under the *Privacy Act 1988* and the *Freedom of Information Act 1982*.
- The **Australian Communications and Media Authority** is responsible for safeguarding the community from harms relating to illegal online gambling activities by enforcing compliance with the *Interactive Gambling Act 2021*.
- The **Classification Board, the Classification Review Board and Ministerially approved classification tools** support the classification of content in Australia under the *Classification (Publications, Films and Computer Games) Act 1995.*
- **The Australian Competition and Consumer Commission** regulates Australian Consumer Law, including online transactions, and coordinates government, law enforcement and the private sector to combat scams through the National Anti-Scam Centre.
- **Uniform defamation laws** were enacted in Australian states and territories in 2005 and 2006, with reforms being progressed through the development and adoption of model defamation provisions. The uniform defamation laws apply to online communications, such as posting defamatory material on social media services.
- eSafety's powers complement the roles of legal advisers and **law enforcement agencies** in addressing online harms. Some online harms, such as menacing, harassing or offending others online, using a carriage service in relation to child abuse material, and sharing or threatening to share, a nude or sexual image or video without the consent of the person shown are crimes in Australia under the Commonwealth Criminal Code.[118]
- In January 2024, amendments to **counter-terrorism legislation**[119] took effect establishing new criminal offences for the public display of prohibited Nazi and terrorist organisation symbols, the public performance of the Nazi salute, and for using a carriage service for violent extremist material.
- The **Australian Framework for Generative Artificial Intelligence in Schools**[120] (the Framework) was released by Education Ministers in December 2023, with implementation by states and territories starting from term 1 2024. The Framework provides guidance for schools in using generative artificial intelligence tools to support teaching and learning.
- A range of Government departments and agencies are involved in protecting Australians from **cyber security threats**, including the Department of Home Affairs, the Department of Defence, the Australian Signals Directorate, and the Attorney-General's Department.
- The Digital Platform Regulators Forum (DP-REG) is an information-sharing and collaboration initiative between Australian independent regulators with a shared goal of ensuring Australia's digital economy is a safe, trusted, fair, innovative and competitive space.

---

[118] Criminal Code, see for example, section 474.17, section 474.22 and section 474.17A.

[119] *Counter-Terrorism Legislation Amendment (Prohibited Hate Symbols and Other Measures) Act 2023.*

[120] Department of Education (2023), Australian Framework for Generative Artificial Intelligence (AI) in Schools, Australian Framework for Generative Artificial Intelligence (AI) in Schools - Department of Education, Australian Government, accessed 26 April 2024.

The Government is also progressing other initiatives to support safer digital spaces for Australians online, including:

- **Artificial intelligence:** On 17 January 2024, the Government released its interim response to the Safe and Responsible AI in Australia consultation. The Government response noted further work led by the Department of Industry, Science and Resources to mitigate potential risks associated with artificial intelligence and support safe and responsible artificial intelligence practices, including considering legislative mechanisms for ensuring mandatory safety guardrails for artificial intelligence in high-risk settings.[121]
- **Privacy Act Review:** In September 2023, the Government released its response to the Privacy Act Review. Reforms will ensure Australia's privacy framework is fit-for-purpose in the digital age, and provide Australians with greater transparency and control over their personal information. These are expected to include new provisions to address the practice of doxxing.[122] Doxxing (or doxing) is 'the intentional online exposure of an individual's identity, private information or personal details without their consent.'[123] A Children's Online Privacy code is being developed to apply to online services that are likely to be accessed by children.
- **Misinformation and disinformation:** The Government is developing new legislation to provide the Australian Communications and Media Authority with powers to combat online misinformation and disinformation. These will include information-gathering and rule-making powers, reserve powers to register and enforce industry codes, and reserve powers to make industry standards. There will also be new measures to improve protections for public debate, freedom of speech and religious expression, improve the transparency and accountability of platforms' decision making, and improve public visibility into the efficacy of platform misinformation and disinformation strategies. The Bill also includes measures to ensure online platforms take meaningful steps to prevent and reduce false, deceptive or misleading content likely to amplify hatred against groups in Australian society. The new legislation will also include consequential amendments to the Online Safety Act.
- **Classification reforms:** On 4 April 2024, public consultation commenced to inform the development of options for the second stage of reforms to the National Classification Scheme. The consultation is seeking views on the scope of the National Classification Scheme, how to ensure classification guidelines are aligned with and responsive to evolving community standards and expectations, and explore options for establishing a fit-for-purpose single national regulator responsible for classification.
- **Online dating:** In September 2023, the Government requested online dating services to develop a code of practice to better protect Australians using their services. The voluntary codes are to be in place by June 2024.
- **Algorithms commitment:** As part of the Australian Government Response to the House of Representatives Select Committee on Social Media and Online Safety report, the Department of Infrastructure, Transport, Regional Development, Communications and the Arts and the Department of Home Affairs are progressing work to understand how algorithms operate on digital platforms, identify potential harm attributed to algorithms on digital platforms, and report to Government on

---

[121] Department of Industry, Science and Resources (2024), Supporting responsible AI: discussion paper, Consultation hub | Supporting responsible AI: discussion paper - Consult hub (industry.gov.au), accessed 26 April 2024.

[122] Attorney-General's Department (2024), Media Conference Transcript - Doxxing and hate speech reforms, Defence Force recruiting, High Court decision, Media Conference – Parliament House | Our ministers – Attorney-General's portfolio (ag.gov.au), accessed 26 April 2024.

[123] eSafety Commissioner (2024), Doxing, Doxing | What is doxing or doxxing? | eSafety Commissioner, accessed 15 February 2024.

possible regulatory reform options. The departments are due to report back to Government shortly.[124]

- **Scams:** The Government is currently exploring options for the development of mandatory industry codes to outline the responsibilities of the private sector, including digital platforms, in relation to scam activity.
- **Dispute resolution processes:** The Government's Response to the Australian Competition and Consumer Commission's Digital Platforms Services Inquiry interim report noted that it will undertake further work to develop dispute resolution requirements so Australians can raise and resolve issues experienced online. The Government has called on industry to develop voluntary internal dispute resolution standards by July 2024.[125]
- **Reforms to address hate speech:** The Government is currently working on legislative reforms to strengthen protections against vilification and hate speech, and will seek to complement legislation in the states and territories.[126]

---

[124] Australian Government (2023), Government Response to Social Media and Online Safety Report, 16, 20230330 - Australian Government response to the Social Media and Online Safety inquiry (infrastructure.gov.au) accessed 26 April 2024.

[125] Australian Government (2023), Government Response to ACCC Digital Platform Services Inquiry, Government Response to ACCC Digital Platform Services Inquiry (treasury.gov.au), accessed 26 April 2024.

[126] Standing Council of Attorneys-General, Communique 23 February 2024, Standing Council of Attorneys-General - 23 February 2024 (ag.gov.au), accessed 26 April 2024.

Appendix 2: International approaches

European Union's *Digital Services Act 2022*

The European Union (EU) adopted a systemic regulatory approach to online safety, covering risks stemming from the design or functioning of a platform and risks pertaining to the use of the platform.

The *Digital Services Act 2022* entered into force in November 2022, with requirements for very large online platforms and search engines commencing on 17 February 2023, with the Digital Services Act becoming directly applicable across the EU and all platforms from 17 February 2024. It aims to establish harmonised rules to ensure a safe, predictable and trusted online environment that facilitates innovation and effectively protects fundamental rights (including consumer protections).

The Digital Services Act applies to online intermediary services (including mere conduit services, caching services and hosting services) with a substantial connection to the EU, irrespective of their location or place of establishment. Regulatory obligations are scaled based on the size and type of intermediary service.

Online platforms or search engines that reach more than 10 per cent of consumers in the EU (approximately 45 million consumers) are designated very large online platforms or very large online search engines by the European Commission and have the greatest regulatory obligations. Every six months online platforms must report their average monthly active service recipients so the European Commission can assess if they should be designated a very large online platform or search engine. Examples of platforms currently designated include Instagram, Facebook, Google Play, Google Maps, X/Twitter, LinkedIn, Wikipedia, Bing and Google Search.

The EU Charter of Fundamental Rights is recognised in the aim of the DSA, to set out 'harmonised rules for a safe, predictable and trusted online environment that facilitates innovation and in which fundamental rights enshrined in the Charter, including the principle of consumer protection, are effectively protected.'
Those fundamental rights include freedom of expression, protection of personal data and rights of the child. Under the EU's Digital Services Act, the largest online platforms and search engines are expected to consider risks to fundamental rights, civic discourse and electoral processes when conducting risk assessments.

Key online safety features include:

- ***Systemic risk assessments and mitigation measures***

  Providers of very large online platforms or search engines must undertake risk assessments of systemic risks stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services. Assessments must include consideration of illegal content, risks to fundamental rights, civic discourse and electoral processes and public security, and risks around gender-based violence, public health, children's wellbeing and serious negative consequences to peoples' physical and mental wellbeing. Assessments must consider the impact of specific systems and practices, including recommender systems, moderation systems and data practices.

  They must also implement reasonable, proportionate and effective mitigation measures to address identified risks.

  Providers of online platforms accessible to minors must put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors, on their service.

- ***Transparency reporting***

  Providers of intermediary services are required to report on content moderation activities undertaken in the reporting period at least once per year. Reports must be publicly available.

  Providers of online platforms must report on the number of disputes submitted to out-of-court settlement bodies, and the number of suspensions imposed on service recipients due to frequently providing manifestly illegal content.

- ***Compliance monitoring***

  Providers of very large online platforms or search engines must be independently audited at their own expense at least once a year. They must also provide regulators with data access on request to enable compliance monitoring.

  Hosting services (including online platforms) must have systems that enable individuals or entities to notify them of suspected illegal content on the service. Services must prioritise notices from trusted flaggers appointed by member states.

  Providers of online platforms are required to suspend the provision of services to service recipients that frequently provide manifestly illegal content (for a reasonable period).

- ***Penalties and enforcement***

  Member states are responsible for investigations and setting infringement penalties, up to a maximum value of 6 per cent of the intermediary service provider's annual worldwide turnover in the preceding financial year.

- ***Investigations and enforcement***

  Member states are expected to sufficiently empower their Digital Services Coordinator and the Commission to conduct investigations. The Digital Services Act also provides member states with powers to order intermediary services to provide information on individual service recipients. Very large online platforms and search engines have additional obligations to provide Digital Services Coordinators or the Commission with access to data necessary to assess compliance (on receipt of a reasoned request and with notice).

- ***Cost recovery***

  The European Commission charges designated very large online platforms and search engines an 'annual supervisory fee' on their designation.

United Kingdom's *Online Safety Act 2023*

The UK's *Online Safety Act 2023* received Royal Assent on 26 October 2023. The new regulatory framework aims to make the use of internet services safer for individuals in the UK.

The UK imposes multiple duties on providers of regulated services to identify, mitigate and manage the risks of harm from illegal content and activity and content and activity that is harmful to children. Details on how service providers can meet their obligations will be placed in secondary legislation, codes and guidelines. The regulator (the Office of Communications, or 'Ofcom') is responsible for drafting the codes and guidelines.

Duties apply in relation to illegal content and activity, and content and activity harmful to children, but also encompass system design (safe by design, child safety design, freedom of expression and privacy protections, and service transparency and accountability).

The duties apply to 'providers of regulated services', including:

- **User-to-user platforms** - where users can upload and share content (for example messages, images, videos, comments) that becomes accessible to others. This includes services such as online discussion forums, social media platforms, dating services and online market places.
- **Search services** - search engines that enable users to search numerous websites and databases
- **Services that provide pornographic content**.[127]

Similar to Australia's legislation, the UK Online Safety Act has extra-territorial application. The UK Online Safety Act applies to providers based outside the UK, if the service has a significant number of UK users, the UK is the target market, or the service can include UK users and there are reasonable grounds to believe that UK individuals are at material risk of significant harm.

The UK also incorporated many fundamental rights outlined in the European Convention on Human Rights into legislation in 1998, including the freedom of expression. The UK's Online Safety Act requires service providers to 'have particular regard to' privacy and user freedom of expression when deciding on and implementing safety measures and policies. User-to-user services with the greatest reach and highest risk have additional obligations which include protecting content of democratic importance, and publishing impact assessments of safety measures and policies on user privacy and freedom of expression within the law.

Key online safety features include:

- **Duties of care**

  The UK Online Safety Act imposes a range of duties of care on regulated services in relation to content and activity on the service. These are scaled based on the service's role and reach. Services likely to be accessed by children, and services that provide pornographic content have additional duties. Ofcom is required to draft industry codes outlining the measures services can take to meet their obligations.

  General duties

  All providers of regulated user-to-user services and search services have duties relating to illegal content, including to conduct risk assessments and take proportionate service design and operational measures to prevent or minimise the risk of users encountering illegal content, and to mitigate and manage the risks of harm to individuals. These include consideration of functionalities, algorithms, user content controls, content prioritisation and staff policies and practices.

  Services must also have systems enabling users or affected persons to easily report illegal content, have relevant complaint systems and processes that include complaints on moderation actions and use of technology in a way not envisaged by the terms of service. Services must also 'have particular regard to' privacy and user freedom of expression when deciding on and implementing safety measures and policies.

  Additional duties

  Additional duties apply to organisations that meet certain thresholds based on size, functionality, reach, and other defined factors. The thresholds will be set out in secondary legislation:
  - Category 1 – reserved for user-to-user services with the highest reach and highest risk functionalities
  - Category 2A – covers the highest reach search services;
  - Category 2B – other services with potentially high-risk functionalities.

---

[127] The UK Online Safety Act 2023 singles out providers of pornographic material, placing on them a standalone duty to ensure that children cannot normally access their services.

Category 1 services have additional duties, including (but not limited to) adult user empowerment, publishing children's risk assessments in their terms of service, protecting content of democratic importance, and publishing impact assessments of safety measures and policies on user privacy and freedom of expression within the law. User empowerment duties encompass content relating to suicide or self-harm, eating disorders or associated behaviours, abuse targeting characteristics (race, religion, sex, sexual orientation, disability or gender reassignment), content inciting hatred against groups of people (or a particular race, religion, sex, sexual orientation, people who have a disability, people who have the characteristic of gender reassignment).

Category 2A services have additional duties, including to publish the most recent illegal content risk assessment, and child risk assessment (if applicable).

All providers of services likely to be accessed by children have children's risk assessment and child safety online duties.

Services that provide pornographic content have duties to use age verification or age estimation (or both) to ensure children are not normally able to encounter the content in relation to the service. The method used must be highly effective at correctly determining whether a particular user is a child. Services are required to keep a written record of these measures and publish a summary statement.

- **Powers to require corrective action**

  Ofcom can review regulated services' compliance with duties and requirements under the OSA and require services to take corrective actions where they believe the service provider's compliance with duties or requirements in the UK Online Safety Act is not adequate. The requirements are enforceable by civil proceedings.

- **Compliance and transparency**

  Ofcom can issue a notice requesting information from regulated and ancillary service or facility providers at any time, for the purpose of exercising or deciding to exercise any online safety functions. This could include requesting risk assessments, risk mitigation measures, documents from engineers regarding new features. Offences apply for failure to comply.

  Ofcom must require Category 1, 2A and 2B services to provide a transparency report each year.

- **Protections for privacy and freedom of expression**

  Duties imposed by service providers seek to ensure user's rights to freedom of expression and privacy. All regulated user to user services and search services have duties relating to freedom of expression and privacy. Category 1 services also have a duty to conduct and publish assessments of the impact of safety measures and policies on users' right to freedom of expression within the law and on the privacy of users.

- **Penalties and enforcement**

  The UK Online Safety Act creates new offences, such as failure to comply with an information notice. Senior managers, parent entities, fellow subsidiaries, and controlling individuals may be liable in certain circumstances. It also provides a range of online harm related offences.

  Ofcom is given wide-ranging enforcement powers around failure to meet duties and requirements (confirmation decisions), including the ability to require corrective actions or to issue fines of up to 10 per cent of annual global turnover or £18 million (whichever is greater). Offences or penalties may apply for failure to comply.

In the most extreme cases, with the agreement of the courts, Ofcom will be able to require payment providers, advertisers, and internet service providers to stop working with a Service, preventing it from generating money or being accessed from the UK (business disruption powers).

- **Investigation and information gathering powers**

  The UK Online Safety Act provides the regulator with powers to require a person to provide information required for the purpose of exercising, or deciding whether to exercise, any of their online safety functions. Additional information gathering powers apply in connection with an investigation into the death of a child, including to obtain information on the child's use of a service, content encountered and interactions with that content. When investigating a suspected contravention, the regulator can require a person to attend an interview. The regulator also has powers to enter, inspect and audit.

- **Cost recovery**

  Ofcom can require the provider of a regulated service to pay an annual fee, calculated by reference to the provider's worldwide revenue over a qualifying period (Part 6). The threshold for fee payment and fee calculation will be determined in Regulations. Penalties may apply for failure to pay the fee.

- **Super complaints**

  The UK Online Safety Act enables approved eligible entities to make a complaint about features and/or conduct of one or more regulated service providers that presents a material risk of causing significant adverse impact to service users or the public (including particular groups), including causing significant harm or affecting the right to freedom of expression. Complaints are only admissible if the matter impacts a particularly large number of service users or members of the public. Processes for addressing super-complaints are yet to be detailed in regulations.

  The UK Online Safety Act does not include schemes to address content removal based on individual complaints (individual support is provided through other UK organisations).

Canada's proposed Online Harms Act

In February 2024, the Government of Canada introduced Bill C-63 which, if enacted, would create a new *Online Harms Act*. The Online Harms Act would have eight purposes, including to promote the online safety of persons in Canada, protect children's physical and mental health, reduce harms caused to persons in Canada as a result of harmful content online, and ensure the operators of social media services are transparent and accountable with respect to their duties under the Act. The purposes balance mitigating the risk of exposure to harmful content online, with respect for freedom of expression and participation in public discourse.

The Online Harms Act would establish a Digital Safety Commission of Canada to administer the Act and a Digital Safety Ombudsperson of Canada to support users of social media services and also advocate for the public interest in relation to online safety systemic issues. The Digital Safety Commission would comprise three to five full time members with a designated Chairperson. Members may be designated as the Vice Chairperson, authorised to exercise or perform the Commission's powers, duties and functions, and be formed into committees to conduct the Commission's work. Both the Commission and Ombudsperson would be supported by a Digital Safety Office.

The Online Harms Act would impose duties on the operators of social media services that exceed a prescribed threshold of users or are designated by regulations. This would include livestreaming and user-uploaded adult content services.

The *Canadian Charter of Rights and Freedoms of 1982* protects fundamental rights, including the freedom of expression. Canada's proposed Online Harms Act lists multiple purposes, including balancing risk mitigation of

exposure to harmful content online with respect for freedom of expression, and the constraint that online harms can place on freedom of expression. The regulator would be required to consider freedom of expression, equality rights, privacy rights, and the needs and perspectives of the Indigenous peoples of Canada when making regulations, or issuing guidelines, codes of conduct, and other documents. There would also be a requirement for the operators of regulated social media platforms to notify a person if content they have posted is made inaccessible, and to reconsider that decision if the person appeals.

- **Duties**

  Under the Online Harms Act, social media services would be subject to three duties: a duty to act responsibly, a duty to protect children, and a duty to make certain content inaccessible.

  Duty to act responsibly

  The duty to act responsibly requires service operators to implement measures that are adequate to mitigate the risk that users of the service will be exposed to harmful content on the service. Seven types of harmful content are targeted: Intimate content communicated without consent, content that sexually victimises a child or revictimizes a survivor, content that induces a child to harm themselves, content used to bully a child, content that foments hatred, content that incites violence, and content that incites violent extremism or terrorism.

  Measures are assessed based on effectiveness, the service size, technical and financial capacity, discrimination protections under the *Canadian Human Rights Act,* and any other factors provided in the regulations.

  Service operators must also provide clear and accessible ways to flag harmful content and block users and a dedicated contact person to hear users' concerns about harmful content on the service.

  Additional duties (including 'to protect children')

  These include:

  - To protect children by implementing specified design features specified in the regulations (such as age appropriate design).
  - To make certain types of content inaccessible to persons in Canada, including content that sexually victimises a child or re-victimises a survivor, and intimate content communicated without consent. The content can be flagged by a user or the service operator. The user who posted the content must be notified, and there must be a mechanism for them to appeal.

- **Powers to require corrective action**

  The Commission can make an order requiring an operator to take (or refrain from taking) any measure to ensure compliance with the Act.

- **Compliance and transparency**

  Service providers must keep records of their compliance with the duties and provide an accessible digital safety plan that addresses the duty to act responsibly. The Commission has broad powers to summons, inspect and hold hearings related to compliance with the Act or certain content complaints.

  The Commission also has powers to enable accredited persons to access data included in digital safety plan inventories for research, education, advocacy or awareness activities related to the purposes of the Act. For research projects, the Commission can order a service operator to give a person access to data referred to in the inventory.

- **Protections for political freedoms**

  Political freedoms are addressed in the purposes of the Online Harms Act and regulatory actions of the Commission. In addition to 'promoting the online safety of persons in Canada,' the eight purposes of the Online Harms Act would include:

  - 'considering that exposure to harmful content online impacts the safety and well-being of persons in Canada, mitigate the risk that persons in Canada will be exposed to harmful content online while respecting their freedom of expression' and
  - 'enable persons in Canada to participate fully in public discourse and exercise their freedom of expression online without being hindered by harmful content.

  The Commission is also required to consider freedom of expression, equality rights, privacy rights, and the needs and perspectives of the Indigenous peoples of Canada when making regulations, or issuing guidelines, codes of conduct, and other documents.

  Digital safety plans

  As part of the duty to act responsibly, regulated service operators are required to submit digital safety plans to the Commission for each service they operate. The digital safety plan outlines risk assessments and measures taken to mitigate the risk end users will be exposed to harmful content on the service, including how the effectiveness of measures is assessed. It must also include additional measures taken to protect children, resources allocated to compliance with the duties (and to automated decision-making), and measures to meet mandatory pornography reporting obligations under government legislation. The digital safety plan must be published on the service in an easy-to-read-format.

  Appeal mechanism for content moderation

  Service operators are required to notify the person who communicated the content where a decision is made to make the content inaccessible. The operator must also reconsider the decision at the request of the person notified.

- **Ombudsperson**

  The Online Harms Act would establish an Ombudsperson with the mandate to provide support to users of regulated services and advocate for the public interest with respect to systemic issues related to online safety. The Ombudsperson's powers would include gathering information, highlighting issues by making information gathered publicly available, and directing users to resources that may address the concerns regarding harmful content.

- **Investigation and information gathering powers**

  Canada's proposed Online Harms Act would provide the regulator with broad powers to investigate complaints and ensure an operator's compliance with the Act. These would include summoning a person to appear and give evidence, and hold hearings (details to be determined in rules). Inspectors would have powers of entry (including by remote access) and examination and powers to require a person to provide information.

- **Cost recovery**

  Cost recovery can be provided for under regulations, for the purpose of recovering all, or a portion of any costs incurred by the Commission, the Ombudsperson, or the Office in relation to the exercise of their powers or the performance of their duties and functions. These can include charges payable by the operator of a regulated service, their calculation and payment, and circumstances in which exemptions would apply based on the operator's ability to pay.

Key features in other jurisdictions

Ireland

Several features appear in other legislative approaches that are not present in Australia's current regulatory approach. For example, while Ireland's *Online Safety and Media Regulation Act 2022* takes many similar approaches to the Act in Australia, there are differences in how online safety codes are used and in the online harms addressed. The Coimisiún na Meán (An Coimisiún) will regulate 'harmful online content' and 'age inappropriate content' through binding online safety codes and safety guidance materials.

Coimisiún na Meán can designate the relevant online services (or categories of services) to which codes apply. When deciding to designate a service An Coimisiún will consider the nature and scale of the service and the levels of risk of exposure to harmful online content when using the service.

The codes will address two types of 'harmful online content':

- offence specific categories of online harm (this is online content comprising information that is criminally prohibited from being published or broadcast); and
- online content which, subject to a risk test, may be considered harmful, including cyberbullying material, content that humiliates another person, content that promotes suicide, self-harm or eating disorders.
    - To meet the relevant risk test the online content must either give rise to a risk to a person's life or to a risk of significant harm to a person's physical or mental health, where the harm is reasonably foreseeable.

'Age-inappropriate online content' is content that is likely to be unsuitable for children (either generally or below a particular age) having regard to their capabilities, development, and their rights and interests. This includes in particular content consisting of pornography or realistic representations of gross or gratuitous violence or acts of cruelty.

The online safety codes may require providers to:

- carry out risk and impact assessments in relation to the availability of harmful online content on their services
- implement measures to minimise the availability of harmful online content and the risks arising from the availability of and exposure to such content
- implement measures to protect users from harmful online content
- implement measures in relation to commercial communications on their services (such as advertising or promotional content) that are appropriate to protect the interests of users, particularly children
- prepare reports for the Commission
- implement measures to put in place mechanisms for handling complaints from service users.

Safety by Design may also be considered for inclusion in a future iteration of Ireland's online safety codes, with a possible requirement for platforms and services to undertake a Safety Impact Assessment.[128] Ireland's first code that will cover video-sharing platforms is expected to be adopted in 2024. An Coimisiún's draft online safety code would require providers of video sharing platform services to publish their methodology for conducting safety impact assessments that are effective in identifying and mitigating specified safety issues, incorporating Safety by Design.[129] Safety issues relate to the physical, mental and moral development of minors, the protection of minors from sexual abuse, and the protection of the general public from racism,

---

[128] Coimisiún na Meán (2023) Call For Inputs: Online Safety, accessed 26 April 2024.
[129] Coimisiún na Meán (2023) Consultation Document: Online Safety, accessed 26 April 2024.

xenophobia and incitement to hatred or violence on any grounds protected under Article 21 of the Charter of Fundamental Rights of the European Union. Protected characteristics include sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or other opinion, membership of a national minority, property, birth, disability, age, or sexual orientation.

Cost recovery

Coimisiún na Meán can impose levies on providers of regulated services in order to recover its expenses and working capital requirements. Four levy models have been proposed for different service types: TV broadcasters; radio broadcasters; video-on-demand providers; and designated online services, including video sharing platform services. The levy for video sharing platform services will be calculated as a fixed amount per monthly active user (mirroring the supervisory fee in the EU Digital Services Act for very large online platforms and search engines). The levy is a contractual debt and can be recovered through the judicial system.

Singapore

Singapore's *Online Safety (Miscellaneous Amendments) Act* enables the regulator to issue directions to block Singapore users' access to egregious content. This includes suicide or self-harm, physical or sexual violence and terrorism, child sexual exploitation, content posing public health risks, and content likely to cause racial and religious disharmony in Singapore. The regulator may also designate regulated online communication services to comply with Codes of Practice to implement systems and processes to mitigate risks to Singapore users from exposure to harmful content and to provide accountability. The Online Safety Code for Designated Social Media Services, designed to enhance user safety and reduce the spread of harmful content, came into effect in 2023. Singapore is also considering a Code of Practice requiring app stores to remove harmful content in the marketplace and an age classification scheme to protect Singaporeans from exposure to harmful content in online games.

Germany

In Germany, the *Youth Protection Act* mandates the use of age verification systems for age-restricted content, including online pornography. A Government body is responsible for assessing the suitability of, and approving, the use of specific age assurance technology providers. This includes the use of identity document scans and biometrics.