Online Safety Codes and Standards

Regulatory Guidance

April 2026

Australian Government | eSafetyCommissioner

eSafety.gov.au

Contents

Overview of this guidance

This guidance is for service providers in sections of the online industry who are regulated by the Online Safety Codes and Standards (**Codes and Standards**). These are the:

- Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) (**Unlawful Material Codes)**)
- Relevant Electronic Services – Class 1A and Class 1B Material Industry Standard 2024 and Designated Internet Services – Class 1A and Class 1B Material Industry Standard 2024 (together, **Unlawful Material Standards**)
- Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) (**Age-Restricted Material Codes**).

The Codes and Standards can be found on the Register of Online Safety Codes and Standards.

This guidance provides information about the Codes and Standards and the functions of eSafety in monitoring and enforcing compliance.

Codes and Standards cover the following industry sections:

- social media services
- relevant electronic services
- designated internet services
- search engine services
- app distribution services
- hosting services
- internet carriage services
- equipment services (including manufacturers, suppliers, and those who maintain and install equipment that is used to access online services).

Part 1: Legal and regulatory framework for the Codes and Standards

eSafety is Australia's independent online safety regulator.

The *Online Safety Act 2021* (Cth) (**the Act**) provides eSafety with legislative powers to promote and improve online safety for Australians. This includes the registration and enforcement of the Online Safety Codes and Standards relating to unlawful and age-restricted material online (often referred to as 'illegal and restricted online content').

Under these Codes and Standards, service providers (participants in specific sections of the online industry[1]) are required to take steps to address the presence of this material on their services for people in Australia who are users of the service (we refer to these as 'end-users'). The aim of the Codes and Standards is to address systemic risks and focus on proactive and systemic change. As such, eSafety's compliance and enforcement activities generally target systemic safety failures, rather than focusing on isolated incidents.

The background to the development of the Codes and Standards can be found on eSafety's website at:

- Background to the Unlawful Material Codes
- Background to the Unlawful Material Standards
- Background to the Age-Restricted Codes

The Codes and Standards are risk-based, with requirements placed on service providers that are proportionate to the risk their service presents with respect to class 1 and 2 material under the National Classification Scheme (for more information see **Table 1**). The requirements in the Codes and Standards are also outcomes-based, setting out the objectives while remaining technology-neutral, and allowing service providers to take different approaches to achieve these outcomes.

References to 'requirement(s)' in this guidance mean 'minimum compliance measures' in each code, and both 'requirements' and 'obligations' in each standards.

[1] Sections of the online industry are specified in Section 135 of the *Online Safety Act 2021* (**the Act**). Section 136 provides that a person is a participant in a section of the online industry if the person is a member of a group that constitutes a section of the online industry.

Requirements under each code and standard are mandatory and enforceable from the commencement date, unless they are specified as optional, or a service provider is exempt based on their risk profile or category.

eSafety can receive complaints and investigate potential breaches of the Codes and Standards.[2] eSafety has a range of enforcement options for non-compliance ranging from administrative resolutions through to civil penalty proceedings with the ability to seek a civil penalty per contravention of up to 30,000 penalty units for individuals and five times this for corporations.[3]

Enforcement is discussed in more detail at **Part 6** of this guidance.

What is the difference between codes and standards?

Codes are developed by industry associations representing service providers in various sections of the online industry, and registered by eSafety if they meet procedural and statutory requirements. The eSafety Commissioner (**the Commissioner**) must be satisfied that an industry-drafted code submitted for registration provides appropriate community safeguards for matters of substantial relevance to the community before registering it as a code.[4]

The Commissioner may determine that a standard applies to service providers in a particular section of the online industry if one or more conditions are met, including if the Commissioner has decided not to register a draft code. eSafety is responsible for developing standards, not industry associations. Standards are legislative instruments that are tabled in the Australian Parliament. They are not technical standards of the kind developed by standards-making bodies.

Requirements under codes and standards are enforceable and compliance is mandatory. A breach of a standard is enforceable through civil penalty proceedings. A breach of a code may result in a direction to comply with the code which, if not complied with, is enforceable through civil penalty proceedings. More information about enforcement is provided in **Part 6**.

[2] Sections 40, 42 of the Act.
[3] Sections 143-144, 146-147 and Part 10 of the Act. The monetary value of 1 penalty unit is $330 (at the date of this regulatory guidance). In addition, the maximum penalty ordered by a Court against a corporation (which can include online service providers) can be five times more than the maximum penalty ordered against an individual.
[4] Section 140 of the Act.

What material is covered by the Codes and Standards?

The Codes and Standards regulate online activities[5] related to class 1 and class 2 material. Class 1 and class 2 material ranges from the most seriously harmful types of content, such as material showing the sexual abuse of children or promoting acts of terrorism, through to material which is inappropriate for children, such as online pornography. eSafety refers broadly to such content as 'illegal and restricted online content.'

Online material regulated under the Codes and Standards includes written, video, audio and/or image-based material, whether it is real or fake (including AI-generated content). Class 1 and class 2 material is defined under the Act by reference to Australia's National Classification Scheme.[6] The definitions in the Act apply to films, publications, computer games and any other material.[7] Additional information on the classification of material is available in eSafety's Online Content Scheme Regulatory Guidance.

Phased development

In a September 2021 Position Paper, eSafety encouraged the online industry to adopt a two-phased approach to the development of codes of standards, prioritising efforts for preventing or mitigating material with the greatest potential for harm in the first instance. The industry associations that drafted the codes followed the recommendation. They also adopted sub-categories for the classes of material.

The first phase covered the Unlawful Material Codes and Standards dealing with class 1A and class 1B material. These were initially referred to as the 'Phase 1 Codes' and the 'Phase 1 Standards'.

The second phase covered the Age-Restricted Material Codes dealing with class 1C and class 2 material. These were initially referred to as the 'Phase 2 Codes'.

[5] Online activities are listed in Section 134 of the Act.
[6] A cooperative arrangement between the Australian Government and state and territory governments for the classification of films, computer games and certain publications. For further information visit the Australian Classification website at www.classification.gov.au.
[7] Other material is material that is not a film, publication or computer game: Sections 106-107 of the Act.

The sub-categories of material and their development phases are outlined in more detail in Table 1.

Table 1: Development phases and subcategories for codes and standards

Development Phase	Sub-category	Material	National Classification Scheme
Phase 1	Class 1A	• Child sexual exploitation material (**CSEM**) – material that is child sexual abuse material[8], that contains exploitative descriptions or depictions of a child, or that promotes or provides instruction of paedophile activity. • Pro-terror material – material that advocates the doing of a terrorist act (including terrorist manifestos). • Extreme crime and violence material – material that describes, depicts, expresses or otherwise deals with matters of extreme crime, cruelty or violence (including sexual violence) **without justification**.[9] For example, murder, suicide, torture and rape. Material that promotes, incites or instructs in matters of extreme crime or violence.	• Class 1 • Refused Classification (RC)
Phase 1	Class 1B	• Crime and violence material – material that describes, depicts, expresses or otherwise deals with matters of crime, cruelty or violence **without justification**. Material that promotes, incites or instructs in matters of crime or violence. • Drug-related material – material that describes, depicts, expresses or otherwise deals with matters of drug misuse or addiction, or provides detailed instruction or promotion, **without justification**.	• Class 1 • Refused Classification (RC)

[8] Child sexual abuse material (**CSAM**), which shows a sexual assault against a child, is a narrower category and can be considered a sub-set of child sexual exploitation material (**CSEM**).

[9] Reference to 'without justification' highlights that the nature of the material must be considered, including its literary, artistic, or educational merit and whether it serves a medical, legal, social or scientific purpose. Section 11 of the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) sets out matters to be taken into account in making a decision on classification.

Development Phase	Sub-category	Material	National Classification Scheme
Phase 2	Class 1C	• Online pornography – material that describes or depicts specific fetish practices or fantasies.	• Class 1 • Refused Classification (RC)
Phase 2	Class 2A	• Online pornography – other sexually explicit material that depicts actual (not simulated) sex between consenting adults.	• Class 2 • X18+ • Category 2 restricted
Phase 2	Class 2B	• Online pornography – material which includes realistically simulated sexual activity between adults. Material which includes high-impact[10] nudity. • Other high-impact material which includes high-impact sex, nudity, violence, drug use, language and themes. 'Themes' includes social issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism.	• Class 2 • R18+ • Category 1 restricted

Which sections of the online industry are regulated by the Codes and Standards?

The Codes and Standards apply to eight sections of the online industry,[11] outlined in **Table 2**.

Table 2: Industry sections and services covered by the Codes and Standards

Industry section	Examples of services (non-exhaustive)
Social media services[12]	• social networks • public media sharing networks • discussion forums • consumer review networks

[10] Impact may be higher where content is detailed, accentuated, or uses special effects, prolonged, repeated frequently, realistic or encourages interactivity.
[11] These sections of the online industry are listed in Section 135 of the Act.
[12] As outlined Part 2 of this guidance, where a service provider offers instant messaging, chat functionality, or any other functionality contained in the definition of relevant electronic services in Section 13A of the Act, that service will likely be required to comply with the Relevant Electronic Services Standard (Unlawful Material). This will affect providers of social media services, regardless of whether they predominantly provide social media functionality.

Industry section	Examples of services (non-exhaustive)
Relevant electronic services	• instant messaging services • Short Message Services and Multimedia Message Services • chat services • online multi-player gaming services • email services • online dating services • enterprise messaging services
Designated internet services	• file storage services managed by end-users in Australia • services that use machine learning models to enable users to generate material (generative AI services) • other websites and apps Note: Unless an online service meets the definition of a social media service or a relevant electronic service, in which case it cannot be Designated Internet Service.
Search engine services	• electronic services designed to collect, organise (index) and/or rank information on the World Wide Web in response to end-user queries and return search results to end-user queries Note: Excludes search functionality within platforms where content or information can only be surfaced from that which has been generated/uploaded/created within the platform itself and not from the World Wide Web more broadly.
App distribution services	• services distributing apps that can be accessed by end-users in Australia (for example app stores/marketplaces) Note: Excludes links to apps and download of apps from third party websites.
Hosting services	• services which host stored material in Australia (for example services with data centres located in Australia)
Internet carriage services	• retail internet service providers (ISPs) that supply internet carriage services (including mobile and broadband) to end-users in Australia Note: Excludes providers of wholesale ISP services, including NBN Co.

Equipment services	• manufacturers, suppliers, maintainers and installers of equipment that is used to access online services[11] such as:
	  ○ mobile phones
	  ○ laptops
	  ○ tablets
	  ○ internet-enabled devices (such as smart TVs and gaming consoles)
	  ○ immersive technologies (such as virtual reality headsets)
	  ○ wi-fi routers
	Note: This section of the online industry includes manufacturers of these devices, as well as businesses and retail outlets that install, sell and/or repair or maintain such devices.

Which codes or standards apply to each section of the online industry?

Each code or standard commences six months after its date of registration by eSafety. See **Part 6** for eSafety's approach to assessing compliance and enforcement.

The full text of each code and standard and its date of registration can be found on the Register of Online Safety Codes and Standards.

Unless specified otherwise, eSafety may start compliance and enforcement actions as soon as the Codes and Standards commence. **Table 3** summarises the codes and standards applicable to industry sections and their commencement dates.

Table 3: Codes and Standards applicable to industry sections and their commencement dates

Industry section	Applicable code or standard	Structure	Commencement date
Social media Services	Social Media Services Online Safety Code (Class 1A and Class 1B Material) (**Social Media Services Code (Unlawful Material)**)	Head Terms + Schedule 1	16 December 2023
	Social Media Services (Core Features) Online Safety Code (Class 1C and Class 2 Material) (**Social Media Services (Core Features) Code (Age-Restricted Material)**)	Head Terms + Schedule 4	9 March 2026
	Social Media Services (Messaging Features) Online Safety Code (Class 1C and Class 2 Material) (**Social Media Services (Messaging Features) Code (Age-Restricted Material)**)	Head Terms + Schedule 4A*	9 March 2026
Relevant electronic services	Relevant Electronic Services – Class 1A and Class 1B Material Standard (**Relevant Electronic Services Standard (Unlawful Material)**)	Standard	22 December 2024
	Relevant Electronic Services Online Safety Code (Class 1C and Class 2 Material) (**Relevant Electronic Services Code (Age-Restricted Material)**)	Head Terms + Schedule 5	9 March 2026
Designated internet services	Designated Internet Services – Class 1A and Class 1B Material Standard (**Designated Internet Services Standard (Unlawful Material)**)	Standard	22 December 2024
	Designated Internet Services Online Safety Code (Class 1C and Class 2 Material) (**Designated Internet Services Code (Age-Restricted Material)**)	Head Terms + Schedule 6	9 March 2026

Industry section	Applicable code or standard	Structure	Commencement date
Internet search engine services	Internet Search Engine Services Online Safety Code (Class 1A and Class 1B Material) (**Search Engine Services Code (Unlawful Material)**)	Head Terms + Schedule 6	12 March 2024
	Internet Search Engine Services Online Safety Code (Class 1C and Class 2 Material (**Search Engine Services Code (Age-Restricted Material)**)	Head Terms + Schedule 3	27 December 2025**
App distribution services	App Distribution Services Online Safety Code (Class 1A and Class 1B Material) (**App Distribution Services Code (Unlawful Material)**)	Head Terms + Schedule 2	16 December 2023
	App Distribution Services Online Safety Code (Class 1C and Class 2 Material) (**App Distribution Services Code** (**Age-Restricted Material)**)	Head Terms + Schedule 7	9 March 2026**
Hosting services	Hosting Services Online Safety Code (Class 1A and Class 1B Material) (**Hosting Services Code (Unlawful Material)**)	Head Terms + Schedule 3	16 December 2023
	Hosting Services Online Safety Code (Class 1C and Class 2 Material) (**Hosting Services Code (Age-Restricted Material)**)	Head Terms + Schedule 1	27 December 2025
Internet carriage services	Internet Carriage Services Online Safety Code (Class 1A and Class 1B Material) (**Internet Carriage Service Code (Unlawful Material)**)	Head Terms + Schedule 4	16 December 2023
	Internet Carriage Services Online Safety Code (Class 1C and Class 2 Material) (**Internet Carriage Service Code (Age-Restricted Material)**)	Head Terms + Schedule 2	27 December 2025
Equipment services	Equipment Online Safety Code (Class 1A and Class 1B Material) (**Equipment Code (Unlawful Material)**)	Head Terms + Schedule 5	16 December 2023
	Equipment Online Safety Code (Class 1C and Class 2 Material) (**Equipment Code (Age-Restricted Material)**)	Head Terms + Schedule 8	9 March 2026

*** For social media services that also have a messaging function, they will be required to treat the Social Media Services (Core Features) Code (Age-Restricted Material) and Social Media Services (Messaging Features) Code (Age-Restricted Material) as a single industry code that will apply to that service.**

**** Age assurance measures in these Codes come into effect 6 months after the commencement date.**

Part 2: Applying the relevant code or standard, and information management

This part provides guidance on applying the codes or standards to each service and its risk profile. This part also provides information about record keeping and the confidentiality of information provided to eSafety.

Which code or standard applies?

Service providers must comply with one Unlawful Material Code or Standard, and one Age-Restricted Material Code, for each separate service they provide.[13] If a service meets the definition of multiple service types under the Act, and could be subject to several codes or standards, the service provider is generally required to comply with the code or standard that reflects the service's predominant purpose.

However, in relation to the Unlawful Material Codes and Standards, if a service is both a relevant electronic service as well as another type of service, then the service provider must comply with the Relevant Electronic Services Standard (Unlawful Material).[14] Further information is provided below under the example for 'Service providers that align with both a social media service and relevant electronic service under the Unlawful Material Codes and Standards.'

Service providers should have regard to the following non-exhaustive factors to assist in distinguishing the services they provide from each other:

- The presence of a separate sign-up process, including terms and conditions, for each service.
- The method(s) by which end-users can access each service (for example, whether via the same website or application).
- The functionality of each service and the level of integration of the functionality between the services (such as, what the service can do).

Where the service provider operates multiple services, it will be required to comply with codes or standards in relation to each service it operates. The code or standard that applies

[13] For social media services that also have a messaging function, they will be required to treat the Social Media Services (Core Features) Code (Age-Restricted Material) and the Social Media Services (Messaging Features) Code (Age-Restricted Material) as a single industry code that will apply to that service.
[14] Section 150 of the Act; Relevant Electronic Services Standard (Unlawful Material) s 5.

to one service a service provider operates may be the same or different to those which apply to another service it operates.

Example 1

A service provider operates an internet carriage service that falls within the Internet Service Provider Codes. That service provider also manufactures and/or supplies equipment that falls within the Equipment Codes.

As a result, the service provider would need to comply with both Internet Service Provider Codes in respect of its internet carriage service and both Equipment Codes in respect of its manufacture and/or supply of equipment.

Example 2

A service provider manufactures and/or supplies equipment that is for use by end-users in Australia and makes available an online messaging service on those devices.

The service provider would need to comply with both Equipment Codes in respect of its manufacture and/or supply of equipment, the Relevant Electronic Services Code (Age-Restricted Material) **and** the Relevant Electronic Services Standard (Unlawful Material) in respect of its online messaging service.

Service providers that align with both a social media service and relevant electronic service under the Unlawful Material Codes and Standards

Many services fulfil multiple purposes and offer several features for users, and meet multiple definitions under the Act. In particular, some social media services are also relevant electronic services. For example, a social media service may offer a range of features that enable online social interaction and allow users to post content as outlined in the Act's definition for a social media service. If that service also includes, for example, instant messaging or chat functionality, then that service is likely to meet the definition of 'relevant electronic service' in the Act and therefore be subject to the Relevant Electronic Services Standard. The presence of other features relating to social media does not affect this categorisation.

This is because Section 5.2 of the Relevant Electronic Services Standard provides that it applies 'to the exclusion of any industry code'. The Act also provides that industry standards prevail over inconsistent industry codes.[15]

More information about service providers that align with both a social media service and relevant electronic service can be found at the web page Frequently asked questions about applying the correct Unlawful Material Standard or Code.

Service providers that align with both a social media service and relevant electronic service under the Age-Restricted Codes

There are two Social Media Services Codes for age-restricted material. For social media services that also have a messaging function, they will be required to treat the Social Media Services (Core Features) Code and the Social Media Services (Messaging Features) Code as a single industry code that will apply to that service.[16] By registering a second Social Media Service Code containing the same messaging requirements as those that apply in the Relevant Electronic Services Code, social media services can still have and comply with obligations tailored to their messaging features, while also complying with the 'core features' Social Media Services Code.

Service providers that meet the definition of a relevant electronic service and social media service are required to comply with the Age-Restricted Material Code (or, in the case of social media services, Codes) that align with the predominant purpose of the service. Service providers are also still required to comply the Relevant Electronic Services Standard (Unlawful Material) which prevails over the Social Media Services Code (Unlawful Material).

For the avoidance of doubt, for example, this may mean that a service which must comply with the Relevant Electronic Services Standard (Unlawful Material) may be required to comply with the Social Media Services (Core Features) and Social Media Services (Messaging Features) Codes (Age-Restricted Material), rather than the Relevant Electronic Services Code (Age-Restricted Material).

Service providers that are age-restricted social media platforms under the Social Media Minimum Age obligation

Age-restricted social media platforms, as defined in Section 63C of the Act, also have obligations under the Codes and Standards. This ensures that there are protections for all users, including for users with age-restricted social media platforms such as 16- to 17-year-

[15] Section 150 of the Act.
[16] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, p 5.

old users. Age-restricted social media platforms will be required to take reasonable steps to prevent users under 16 years from having accounts on their services.

Further information is provided in **Part 5** and **Appendix F**.

Assessing risk and categorisation under the Codes and Standards

Some of the Codes and Standards differentiate between the levels of risk that different kinds of services can pose to end-users in Australia in relation to class 1 and class 2 material. This ensures requirements in each of those codes and standards are proportionate to, and appropriate for, the risk that class 1 or class 2 material will be generated by, accessed by, or distributed to end-users in Australia, or stored on the service.

Where relevant, 'pre-assessed' and 'defined' categories of services have their risk profiles pre-determined in the relevant code or standard. This recognises the commonality of the features and/or risks on these services. These service providers do not need to conduct their own risk assessments for these services. Pre-assessed and defined categories of services that are designated internet services or relevant electronic services under the Unlawful Material Standards are set out in **Appendix A**.

Providers of services that do not fall within the pre-assessed or defined categories are required either to conduct their own risk assessment for each service or default to assigning it a Tier 1 (high) risk profile.

If a service provider opts to self-assign a service they provide as a Tier 1 (high) risk profile under an applicable code (not standard), they are obligated to notify eSafety. More information on this requirement can be found in **Part 3**.

If a risk assessment indicates that the service may be in-between risk tiers, the **service provider should assign a higher risk profile to that service**. This reflects the fact that while some risks on a service may be lower, it only requires one higher risk feature or attribute to increase harm. **Table 4** sets out risk assessment requirements for each section of industry.

Table 4: Risk assessment requirements for each section of industry

Section of industry		Risk assessment or categorisation required	Risk tiers, pre-assessed or defined categories
App Distribution Services	Code (Unlawful Material)	✗	N/A
	Code (Age-Restricted Material)	✓ Clause 7.1(c)(ii)	N/A
Designated Internet Services (DIS)	Standard (Unlawful Material)	✓ Sections 6(2), 7-8	• Tier 1 (including a high impact DIS) • Tier 2 • Tier 3 (including pre-assessed classified DIS; pre-assessed general purpose DIS; enterprise DIS) • an end-user managed hosting service • high impact generative AI DIS • a model distribution platform
	Code (Age-Restricted Material)	✓ Clause 4.2 (for certain services)	• Tier 1 (including a high impact class 2 DIS) • Tier 2 • Tier 3 • High impact generative AI DIS • End-user managed hosting service • Classified DIS • Model distribution platform • General Purpose DIS
Equipment Services	Code (Unlawful Material)	✓ Clause 5	• Interactive (Tier 1) device (including children's interactive device) • Secondary (Tier 2) device • Non-interactive (Tier 3) device • Gaming device • Operating system provider
	Code (Age-Restricted Material)	✓ Clauses 5-6	• Interactive (Tier 1) devices • Secondary (Tier 2) devices • Non-interactive (Tier 3) devices • Other interactive devices
Social Media Services	Code (Unlawful Material)	✓ Clauses 4-5	• Tier 1 • Tier 2 • Tier 3
Social Media Services (Core Features)	Code (Age-Restricted Material)	✓ Clauses 4-5	• Tier 1 • Tier 2 • Tier 3

Section of industry		Risk assessment or categorisation required	Risk tiers, pre-assessed or defined categories
Social Media Services (Messaging Features)	Code (Age-Restricted Material)	✗	N/A
Relevant Electronic Services (RES)	Standard (Unlawful Material)	✓ Sections 7-8	• Tier 1 • Tier 2 • Tier 3 • Communication RES • Dating services • Enterprise RES • Gaming services with communications functionality • Gaming services with limited communications functionality • Telephony RES
	Code (Age-Restricted Material)	✓ Clauses 4-5	• Tier 1 • Tier 2 • Tier 3 • Other communication RES • Closed communication RES • Dating service • Enterprise RES • Gaming service with communications functionality • Gaming service with limited communications functionality
Hosting Services	Code (Unlawful Material)	✗	N/A
	Code (Age-Restricted Material)	✗	N/A
Internet Carriage Services	Code (Unlawful Material)	✗	N/A
	Code (Age-Restricted Material)	✗	N/A
Search Engine Services	Code (Unlawful Material)	✗	N/A
	Code (Age-Restricted Material)	✓ Clause 5 (for assessing what constitutes appropriate action)	N/A

Assessing risk and categorisation for the Designated Internet Services and Relevant Electronic Services Standards (Unlawful Material)

For services that do not fall into pre-assessed or defined categories, eSafety recommends that service providers use the risk profiles outlined in **Appendix B** as a guide when conducting a risk assessment under Part 3 of these standards. The tables are designed to support interpretation of the risk methodology matters which must be considered under Section 8(5) for both types of services.. The risk assessment methodology does not provide any weighting to the matters that must be considered.

Depending on the nature of a service and the context it operates in, providers are likely to have additional risk factors to consider as part of their methodology. Some risk factors may not be applicable to a service. Service providers should consider only relevant risk factors.

Noting that each service is different, this guide offers a sliding scale of potential risk indicators which providers can apply as relevant to their services.

Risk profile notifications where highest tier is automatically self-assigned

The Social Media Services Code (Unlawful Material) and Social Media Services (Core Features) Code (Age-Restricted Material) require a service provider to proactively notify eSafety of the risk profile or category if it has self-assigned by default the highest risk tier to its service.[17]

The required notification must have been made to eSafety on or before the date the applicable code or standard comes or came into effect.[18]

[17] Social Media Services Code (Unlawful Material), cl 4.1; Social Media Services (Core Features) Code (Age-Restricted Material), cl 4.

[18] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 5.2(a)(ii); Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.2(a)(ii).

eSafety may request information from a service provider about their risk assessment

Under the Codes and Standards, eSafety can still require information about risk profiles from a service provider even if the highest risk tier has been self-assigned by default.[19]

For other services that have not automatically assumed the highest risk tier, or those regulated under the Unlawful Material Standards, service providers do not need to proactively notify eSafety of their risk profiles or categories. However, where a service provider is required to conduct a risk assessment, eSafety can seek this information from a service provider by requiring documents outlining the risk profile determined by a service provider and details of their risk assessment, as well as reasons for assigning a particular risk profile or category.[20]

Assessing risk when making a material change to a service

Table 5 outlines the requirements for service providers to assess risk when making a material change on any of their services. Before making material changes, providers of certain services must also assess the kinds of features and settings that could be implemented on the service to minimise the risk that class 1 and 2 material will be accessed by or distributed to end-users, or which would be stored on the service. This applies irrespective of whether the service provider is captured under a defined or pre-assessed category and would be ordinarily exempt from other requirements to undertake a risk assessment. This ensures that services which undergo substantial changes maintain appropriate categorisation under the Codes and Standards.

[19] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 5.2; Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.2.

[20] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 5.2; Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.2; Designated Internet Services Standard (Unlawful Material), s 31; Relevant Electronic Services Standard (Unlawful Material), s 32.

Table 5: Requirement for industry participants to assess risk when making a material change on their service

Section of industry		Risk assessment required when making a material change
App Distribution Services	Code (Unlawful Material)	✗
	Code (Age-Restricted Material)	✓ Compliance measure 7.13
Designated Internet Services (DIS)	Code (Unlawful Material)	✓ Compliance measure 24 (Tier 1; Tier 2; End-User Managed Hosting Service; High Impact Generative AI DIS)
	Code (Age-Restricted Material)	✓ Compliance measure 7.21 (Tier 1; Tier 2; Tier 3) ✓ Compliance measures 10.17 and 10.26 (Tier 1 High Impact Generative AI DIS)
Equipment	Code (Unlawful Material)	✗
	Code (Age-Restricted Material)	✓ Compliance measure 23 (Operating Service Providers of interactive (Tier 1) Devices)
Social Media Services	Code (Unlawful Material)	✓ Clause 4.4
Social Media Services (Core Features)	Code (Age-Restricted Material)	✓ Compliance measure 10.10 (Tier 1 AI Companion Chatbots)
Social Media Services (Messaging Features)	Code (Age-Restricted Material)	✓ Compliance measure 10.7
Relevant Electronic Services (RES)	Standard (Unlawful Material)	✓ Compliance measure 18 (Pre-assessed, Tier 1, Tier 2)
	Code (Age-Restricted Material)	Compliance measure 8.7 (Closed communication RES) Compliance measure 9.7 (Other communication RES) Compliance measure 10.8 (Dating Services, unless age assurance and access control measures have been implemented) Compliance measure 11.7 (Gaming Services with communications functionality) Compliance measure 15.6 (Tier 1) Compliance measure 16.8 (Tier 1 AI Companion Chatbot Features)

Section of industry		Risk assessment required when making a material change
Search Engine Services	Code (Unlawful Material)	✗
	Code (Age-Restricted Material)	✓ Compliance measure 20
Internet Carriage Services	Code (Unlawful Material)	✗
	Code (Age-Restricted Material)	✗
Hosting Services	Code (Unlawful Material)	✗
	Code (Age-Restricted Material)	✓ Compliance measure 5

Information management

Record keeping requirement

Service providers must keep records of the compliance measures they have adopted for the previous two years.[21]

This information will help eSafety to assess whether service providers are fulfilling their obligations under the Codes and Standards.

Information should be stored in a format that allows records to be retrieved and provided to eSafety if required. Service providers should retain an appropriate amount of detail in these records to assist eSafety to assess compliance.

Confidentiality of information provided to eSafety in reports and other compliance items

Generally, eSafety does not intend to publish compliance reports or confidential information provided by service providers. This does not however limit the Commissioner's ability to exercise their functions under the Act.

[21] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 7.2(b), Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 7.2(b); Relevant Electronic Services Standard (Unlawful Material) s 39(3); Designated Internet Services Standard (Unlawful Material) s 38(3).

The Head Terms to the Codes and Standards outline that if a service provider identifies material in a compliance report as confidential information, eSafety is expected to maintain such material in confidence.[22]

eSafety considers that confidential information includes, but is not limited to:

- information that is commercial-in-confidence (including trade secrets)
- other business information that would be unreasonable to publish
- information that could affect law enforcement and public safety
- personally identifiable information.

However, there may be circumstances in which the Act, or another Australian law, requires or authorises eSafety to disclose this material.

The key purpose of the compliance reports required under the Codes and Standards is to assist eSafety to determine compliance with the applicable Code or Standard and identify whether investigation and/or enforcement is appropriate and necessary. eSafety does not intend to publish compliance reports required under the Codes and Standards. However, the information provided in a compliance report may be relevant to the exercise of statutory powers and functions by eSafety.

eSafety can also be required to produce material in certain circumstances including:

- in response to a request under the *Freedom of Information Act 1982* (Cth)
- at a court's direction or in performance of its duties in court proceedings
- in response to a Minister, house of parliament or another government agency's power to obtain information.

The codes also allow for service providers to refer to information provided under existing voluntary reporting, already provided by the service provider in relation to a different applicable code or standard, or another reporting requirement under the Act.[23] This may include publicly available information or information provided in response to a notice in connection with the Basic Online Safety Expectations (discussed in **Part 5** of this guidance). The purpose of this is to reduce the regulatory burden on service providers and potential duplication.

---

[22] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 7.3(b), Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 7.3(b).

[23] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 7.3(d)-(e), Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 7.3(d)-(e).

Part 3: Complying with communication and reporting requirements under the Codes and Standards

This part provides guidance on the steps that service providers can take to ensure compliance with requirements to communicate with or report to eSafety proactively or on request.

This part does not cover requirements exhaustively. Individual codes and standards and, where relevant, Head Terms to the Codes, should be referred to for a comprehensive understanding of these requirements.

eSafety expects service providers to communicate with eSafety in a timely, appropriate and collaborative manner. Service providers can provide some of the communications and reporting required by the Codes and Standards through the Industry Compliance Portal. More information about the portal can be found on eSafety's website, including how to request access. All other communications can be made by contacting eSafety at codes@eSafety.gov.au.

eSafety's systems will securely store information provided as part of these communications and reporting.

Requirement to update eSafety on relevant changes to service functionality

Some codes and both standards require service providers to provide updates to eSafety on significant new features or changes to service functionality that may have a material effect on increasing the risk of class 1 and class 2 material on the service accessed by end-users in Australia. These are set out in **Table 6**.

Requirements across the Codes and Standards vary in terms of when a service provider is required to notify eSafety of a change to service feature or functionality, and in what format.

Generally, eSafety considers that it would be good practice for service providers to notify eSafety **in advance** of a functionality change, or at least **within two weeks** of implementing a new feature or functionality change. Under the Unlawful Material standards, certain service providers are required to notify eSafety as soon as practicable after making the decision to implement the new feature or function.

In some circumstances, it may be appropriate for a service provider to provide an update through a compliance report when one is required to be given to eSafety, rather than through a separate notification. However, if there is an extended period of time, such as several months, between the implementation of the feature and the provision of the compliance report, we expect that eSafety will be notified of the new feature separately to the compliance report.

Table 6: Notification requirements for changes to service feature or functionality under codes and standards

Code or Standard	Notification required when making change
Social Media Services Code (Unlawful Material)	✓ Compliance measure 19 (Tier 1)
Social Media Services (Core Features) Code (Age-Restricted Material)	✓ Compliance measure 9.10 (all that allow online-pornography, self-harm material or high-impact violence material) ✓ Compliance measure 10.7 (AI Companion Chatbot Features)
Social Media Services (Messaging Features) Code (Age-Restricted Material)	✓ Compliance measure 10.6
App Distribution Services Code (Unlawful Material)	✓ Compliance measure 6
App Distribution Services Code (Age-Restricted Material)	✓ Compliance measure 12
Equipment Code (Unlawful Material)	✓ Compliance measure 4
Equipment Code (Age-Restricted Material)	✓ Compliance measure 22 (Manufacturers of interactive (Tier 1) devices, secondary (Tier 2) devices, other interactive devices; Operating System Providers)
Search Engine Services Code (Unlawful Material)	✓ Compliance measure 8
Search Engine Services Code (Age-Restricted Material)	✓ Compliance measure 21
Relevant Electronic Services Standard (Unlawful Material)	✓ Compliance measure 36 (Pre-assessed; Tier 1)

Code or Standard	Notification required when making change
Relevant Electronic Services (RES) Code (Age-Restricted Material)	✓ Compliance measure 8.6 (Closed Communication RES) ✓ Compliance measure 9.7 (Other Communication RES) ✓ Compliance measure 10.7 (Dating Services) ✓ Compliance measure 11.6 (Gaming Services with Communications Functionality) ✓ Compliance measure 15.12 (Tier 1) ✓ Compliance measure 16.6 (Tier 1 and Tier 2 AI Companion Chatbot Features)
Designated Internet Services (DIS) Standard (Unlawful Material)	✓ Compliance measure 36 (Tier 1; Tier 2; End-User Managed Hosting Service; High Impact Generative AI DIS; Model Distribution Platform)
Designated Internet Services (DIS) Code (Age-Restricted Material)	✓ Compliance measure 7.13 (Tier 1 for relevant high-risk material) ✓ Compliance measure 7.14 (Tier 2 for online pornography and/or self-harm material) ✓ Compliance measure 7.23 (Tier 1; Tier 2; Tier 3 for online pornography and/or self-harm material) ✓ Compliance measure 10.11 (Tier 1 and Tier 2 for generative AI restricted category material) ✓ Compliance measure 10.25 (Tier 1 and Tier 2 for online pornography and/or self-harm material)
Internet Carriage Services Code (Unlawful Material)	✗
Internet Carriage Services Code (Age-Restricted Material)	✗
Hosting Services Code (Unlawful Material)	✗
Hosting Services Code (Age-Restricted Material)	✗

Requirement to refer unresolved complaints to eSafety

Certain service providers are required to refer to eSafety any complaints made about their non-compliance with the applicable code or standard that they have not been able to resolve. This requirement is triggered where a service provider becomes aware that the complainant is dissatisfied with the way in which the report or complaint was dealt with or the outcome of the report or complaint. This requirement is summarised in **Table 7**.

Table 7: Requirements to refer unresolved complaints to eSafety under Online Safety Codes and Standards

Code or Standard	Requirement to refer unresolved complaints about non-compliance
Social Media Services Code (Unlawful Material)	✓ Compliance measure 18 (Tier 1)
Social Media Services (Core Features) Code (Age-Restricted Material)	✓ Compliance measure 9.9 (All social media services that allow class 1C and class 2 material; Tier 1 and Tier 2 social media services that do not allow class 1C and class 2 material)
Social Media Services (Messaging Features) Code (Age-Restricted Material)	✓ Compliance measure 10.17
App Distribution Services Code (Unlawful Material)	×
App Distribution Services Code (Age-Restricted Material)	✓ Compliance measure 9
Equipment Code (Unlawful Material)	×
Equipment Code (Age-Restricted Material)	✓ Compliance measure 18 (Manufacturers and Operating System Providers of interactive (Tier 1) devices)
Search Engine Services Code (Unlawful Material)	✓ Compliance measure 7
Search Engine Services Code (Age-Restricted Material)	✓ Compliance measure 18
Relevant Electronic Services Standard (Unlawful Material)	✓ Compliance measure 31 (Pre-assessed; Tier 1)

Code or Standard	Requirement to refer unresolved complaints about non-compliance
Relevant Electronic Services (RES) Code (Age-Restricted Material)	✓ Compliance measure 8.17 (Closed Communication RES) ✓ Compliance measure 9.17 (Other communication RES) ✓ Compliance measure 10.18 (Dating services) ✓ Compliance measure 11.17 (Gaming Services with communications functionality) ✓ Compliance measure 15.17 (Tier 1; Tier 2)
Designated Internet Services (DIS) Standard (Unlawful Material)	✓ Compliance measure 30 (End-User Managed Hosting Services; High Impact Generative AI DIS; Tier 1)
Designated Internet Services (DIS) Code (Age-Restricted Material)	✓ Compliance measure 7.11 (Tier 1) ✓ Compliance measure 8.10 (End-user Managed Hosting Services) ✓ Compliance measure 10.8 (Tier 1 for generative AI restricted category material)
Internet Carriage Services Code (Unlawful Material)	×
Internet Carriage Services Code (Age-Restricted Material)	×
Hosting Services Code (Unlawful Material)	×
Hosting Services Code (Age-Restricted Material)	×

A service provider will **not** satisfy this obligation by merely referring a complainant to eSafety's codes and standards complaint form.

Service providers can notify eSafety by submitting the information through the Codes and Standards Industry Compliance Portal, outlining:

- the details of the complaint, including the service provider's internal reference number
- to the extent that the service provider is aware, the reason that the complainant is dissatisfied.

This should be done as soon as reasonably practicable after becoming aware that a complaint is unresolved.

More information about the portal can be found on eSafety's website. Service providers can request a link to the portal by contacting eSafety at codes@eSafety.gov.au.

The service provider should advise the complainant that details of their complaint have been provided to eSafety to assist with our monitoring and assessment of the Codes and Standards. The service provider should also advise the complainant that if they wish to provide further information about their complaint to eSafety, they can do so through the codes and standards complaint form (including the service provider's internal reference number).

Requirement for reporting to eSafety on compliance

Service providers are required to report to eSafety on their compliance with an applicable Code annually, or otherwise when required for the applicable Code or Standard by eSafety.

Compliance reports are generally required to specify:

- the steps the service provider has taken to comply with the applicable code or standard and why these steps were appropriate
- the number of complaints made to the service provider about the service provider's non-compliance with the code or standard
- depending on the type of service, other information like the average monthly number of active end-users, the volume of child sexual exploitation and pro-terror material identified on the service (Unlawful Material Codes and Standards), and details of the actions taken with regards to the identified material.[24]

eSafety's Industry codes and standards compliance page contains practical guidance on how to access applicable forms or templates.

Annual reporting

Service providers listed in **Table 8** are required to report on an annual basis.

[24] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, cl 7.3. Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 7.3; Relevant Electronic Services Standard (Unlawful Material) s 36; Designated Internet Services Standard (Unlawful Material) s 36. See also each applicable Schedule for the information required under code compliance reports.

Table 8: Annual reporting schedule

Code or Standard	Services	Reporting dates
Social Media Services Code (Unlawful Material) **Compliance measure 32**	Tier 1 social media services	**15 February each year,** covering a 12-month reporting period ending 15 December the previous year.
Equipment Code (Unlawful Material) **Compliance measure 13**	Manufacturers of interactive (Tier 1) devices and Operating System Providers	**15 February each year,** covering a 12-month reporting period ending 15 December the previous year.
Social Media Services (Core Features) Code (Age-Restricted Material)* **Compliance measure 7.4**	Social media services that allow online pornography, self-harm material and high-impact violence material	**8 May each year,** covering a 12-month reporting period ending 8 March the same year.**
*** Includes reporting under the Social Media Services (Messaging Features) Code (Age-Restricted Material)** **** For compliance reports due 8 May 2027, the applicable reporting period is 9 March 2026 – 8 March 2027.**		

If a compliance report is not provided to eSafety as required, or the report suggests non-compliance with the applicable code or standard or does not provide sufficient detail, eSafety may commence an investigation and/or, in the case of a code, issue a service provider with a written direction to comply with the code.[25]

Other reporting

Service providers that are not required to report annually may be required, by written notice from eSafety, to provide the Commissioner a compliance report for the most recent calendar year. Where a compliance report is required to be submitted on eSafety's request, the service provider must submit their compliance report **within 2 months.**

eSafety requests for reporting on technical feasibility and reasonable practicability (Unlawful Material Standards and Age-Restricted Material Codes)

This reporting requirement applies to the Unlawful Material Standards and the Age-Restricted Material Codes.

[25] Section 143 of the Act.

Where a service provider is required to implement systems, processes and technologies, eSafety can request a report that describes:

- the cases in which it was not, or would not, be technically feasible or reasonably practicable for a service provider to implement systems or technologies of a particular kind to comply with its obligations, including other obligations which refer to technical feasibility or reasonably practicability
- the systems or technologies that were or are available but were not, or would not be, implemented to comply because to do so would introduce a systemic weakness or systemic vulnerability, or would require an end-to-end encrypted service to implement a new form of decryption
- the alternative action taken to comply where a service provider is required to implement appropriate alternative action.[26]

For more information on terminology related to 'technical feasibility' and 'reasonable practicability' see **Appendix C**.

Requirement to notify eSafety of app removals (App Distribution Services Code (Unlawful Material))

The App Distribution Services Code (Unlawful Material) requires app distribution service providers to notify eSafety if they remove a third-party app from their service, where the removal relates to the availability of class 1A material.[27] This notification must be made in writing and as soon as reasonably practical.

What is reasonably practical will depend on the circumstances of the particular case. eSafety considers that 24 hours will usually be an appropriate period to notify eSafety. This supports the purpose of the compliance measure, which is to ensure that appropriate and timely action can be taken in relation to the same app's availability on other app distribution services.

App removal notifications are not required under the App Distribution Services Code (Age-Restricted Material).

[26] See, for example, the Relevant Electronic Services Standard (Unlawful Material), s 33.
[27] App Distribution Services Code (Unlawful Material), compliance measure 5.

Part 4: How eSafety can assist service providers

Service providers may seek guidance and information from eSafety if they are unsure which codes or standards apply to them and what steps they should take to meet compliance obligations and measures.

The guidance or information that eSafety will be able to provide in response to such requests will be general in nature. eSafety can provide further information on the interpretation of a provision in an industry code or standard or the Act, but is unable to provide legal advice as to how that provision applies to a specific set of circumstances.

Where service providers are concerned about their legal position with respect to compliance with the Codes and Standards, they should seek their own legal advice.

Service providers can contact eSafety with general enquiries about the various codes and standards through the Codes and Standards Industry Compliance Portal. More information about the portal can be found on eSafety's website, including how to request access.

eSafety will also engage with service providers and industry associations – both informally and during compliance assurance activities – to understand the experiences of service providers during implementation of the Codes and Standards.

eSafety will update its guidance on how to comply with the Codes and Standards if compliance and enforcement issues identified require additional guidance.

Part 5: How do the Codes and Standards interact with other regulatory requirements?

eSafety has a range of legislative functions and powers to regulate harmful online content and activity, including powers to issue removal notices and investigate breaches of service providers' regulatory requirements. Some of these functions and powers interact with regulatory requirements under the various codes and standards. This section outlines how the Codes and Standards may interact with other regulatory requirements in the Act.

Basic Online Safety Expectations

The Basic Online Safety Expectations (**Expectations**) set out the steps the Australian Government expects should be taken by providers of designated internet services, relevant electronic services and social media services to keep end-users in Australia safe online. The Expectations are set out in a determination from the Minister for Communications.[28]

Compliance with the Expectations is not enforceable. However, eSafety has powers under the Act to obtain information from the applicable service providers, on a periodic or non-periodic basis, about the steps they are taking to comply with the Expectations. eSafety can also publish statements about whether service providers have or have not complied with the Expectations and summaries of the information received in response to notices. The aim is to increase the transparency and accountability of service providers, thereby helping to incentivise and improve safety standards.[29]

Interaction with the Codes and Standards

Requirements in the Codes and Standards are more specific and prescriptive than those in the Expectations.

Steps taken to meet Expectations that relate to class 1 and class 2 material are applicable for many requirements under the Codes and Standards. However, compliance with each requirement in the Codes and Standards will be assessed on its own merit.

[28] For the complete Expectations, see *Online Safety (Basic Online Safety Expectations) Determination 2022* (23 January 2022) and associated Explanatory Statement.
[29] See generally Part 4 of the Act. A failure to comply with a reporting notice to the extent that a person is able can attract a civil penalty (up to 500 penalty units) in addition to other enforcement action: Section 50 of the Act.

Similarly, given the breadth of the Expectations, additional steps beyond those set out in the Codes and Standards may be required to meet the applicable Expectations.

eSafety may use information about a service provider's compliance with the Expectations or information published in a transparency report to determine whether to commence or inform an investigation about non-compliance with a code or standard.

eSafety's Regulatory Guidance – Basic Online Safety Expectations contains additional information about the Expectations and highlights where the Expectations may overlap with requirements under Codes and Standards.

Restricted Access System

A Restricted Access System is an access-control system that meets the requirements under the *Online Safety (Restricted Access Systems) Declaration 2022* (Cth) (**RAS Declaration**). This sets out the minimum requirements for access-control systems used by social media services, relevant electronic services and designated internet services provided from Australia. The RAS Declaration's primary aim is to restrict children's access to R18+ online content (a subset of class 2 material), upon receiving a notice from eSafety.

Rather than mandating specific technologies or processes, the RAS Declaration states that an access-control system must:

- require an application be made by a person to access the relevant material, declaring they are at least 18 years of age
- incorporate reasonable steps to confirm an applicant is at least 18 years of age
- give warnings about the nature of the material and safety information about how a parent or guardian may control access to the material
- limit access to the material unless certain steps are followed.

Interaction with the Codes and Standards

The RAS Declaration is applicable to social media services, relevant electronic services and designated internet services provided from Australia. The Codes and Standards have a broader remit and are applicable to all eight sections of the online industry that provide services to end-users in Australia.

Both the RAS and the Age-Restricted Material Codes require services to take steps to confirm if an end-user is 18 years of age and apply access control measures. Further information about what steps are required under the Age-Restricted Material Codes can be found in **Appendix F**.

Online Content Scheme

The Online Content Scheme in Part 9 of the Act sets out the legislative framework for the Codes and Standards. The Online Content Scheme also gives eSafety a range of other powers to deal with class 1 and class 2 material, including removal or restriction of this material.

Interaction with the Codes and Standards

The Codes and Standards deal with class 1 and class 2 material on online services at a **systemic level** while the other powers under the Online Content Scheme relate to **specific identified examples** of class 1 and class 2 material.

The powers under the Online Content Scheme complement the requirements under the Codes and Standards.

Providers of social media services, relevant electronic services, designated internet services, hosting services, app distribution services and search engine services must comply with any applicable notices issued by eSafety in relation to specific content and must also comply with the applicable code or standard.

Under the Online Content Scheme, eSafety may:

- give a removal notice to providers of social media services, relevant electronic services, designated internet services or hosting services to take all reasonable steps to remove class 1 or class 2 material within 24 hours, or a longer timeframe specified by eSafety[30]
- give a remedial notice to providers of social media services, relevant electronic services, designated internet services or hosting services from Australia to take all reasonable steps to remove class 2 material or to ensure that access to the material is subject to a restricted access system within 24 hours, or a longer timeframe specified by eSafety[31]

[30] Sections 109-110, 114-115 of the Act.
[31] Sections 119-120 of the Act.

- give a written notice to an app distribution service requiring it to cease enabling the download of a particular app when certain requirements under the Act are met[32]
- give a link deletion notice to providers of search engine services requiring the services to stop providing a link that enables access to class 1 material within 24 hours, or a longer timeframe specified by eSafety, when certain requirements under the Act are met.[33]

More information about the Online Content Scheme can be found in our Regulatory Guidance – Online Content Scheme.

Abhorrent violent conduct powers

The Act includes powers which allow eSafety to request or require an internet service provider to block material that promotes, incites, instructs in or depicts 'abhorrent violent conduct' in certain circumstances.[34]

Interaction with the Codes and Standards

The Internet Carriage Services Code (Unlawful Material) operates alongside and complements the abhorrent violent conduct powers and related Online Crisis Protocol. The Internet Carriage Services Code (Unlawful Material) requires an internet service provider to become a signatory to the Online Crisis Protocol on eSafety's request.[35]

More information about abhorrent violent conduct powers can be found in eSafety's Abhorrent Violent Conduct Powers - Regulatory Guidance.

Safety by Design

Safety by Design is an eSafety initiative consisting of a set of principles and assessment tools that position user safety as a fundamental design consideration for online platforms and services.[36] The initiative also includes six foundation modules that provide guidance on

[32] Section 128 of the Act
[33] Section 124 of the Act.
[34] Part 8 of the Act. See also sections 95, 99 of the Act. A protocol developed by eSafety, Australian Internet Service Providers and the Australian Telecommunications Alliance (the industry body responsible for drafting the Internet Service Provider Code, previously known as Communications Alliance) setting out the administrative procedures required to notify Internet Service Providers of a potential online crisis event.
[35] Internet Carriage Service Code (Unlawful Material), compliance measure 3.
[36] One of eSafety's functions under the Act is to formulate written guidelines or statements recommending best practices for promoting and maintaining online safety for Australians: section 27(1)(p)-(q) of the Act.

operationalising the principles, resources for investors and financial entities, and engagement with the tertiary education sector.

The Safety by Design assessment tools and modules are intended to provide both a safety health check and a learning resource that helps companies continually improve online safety. The assessment tools take service providers through sets of targeted multiple-choice questions, as well as information that is relevant to the overarching stream they select.[37] The multiple-choice questions ask service providers about the systems, processes and practices that are in place at their company. The responses generate a tailored report that identifies opportunities to improve user safety.

Interaction with the Codes and Standards

Safety by Design principles and tools, although voluntary, can be used by service providers as a way to support compliance with the Codes and Standards. In particular, Safety by Design tools are referred to in the:

- Social Media Services Code (Unlawful Material) as a way to comply with compliance measures 5 and 13
- Social Media Services (Core Features) Code (Age-Restricted Material) as a way to comply with compliance measure 10.2
- Search Engine Services Code (Unlawful Material) as a way to comply with compliance measure 4
- Designated Internet Services Code (Age-Restricted Material) as a way to comply with compliance measure 10.2
- Relevant Electronic Services Code (Age-Restricted Material) as a way to comply with compliance measure 16.2

As well, Safety by Design principles inform and underpin many compliance measures and requirements under the Codes and Standards.

These tools and their foundational principles provide service providers with realistic, actionable and achievable measures to help safeguard users from online risks and harms. Service providers can use the principles and tools to guide them as they incorporate, assess and enhance user safety for their platforms and products.

[37] Streams include (1) Founder CEO/Director/Founder or (2) Product/Policy/Project Owner or Manager.

More information on Safety by Design, including its principles and tools can be found on eSafety's website.

Social Media Age Restrictions

In December 2024, the Parliament of Australia enacted the *Online Safety Amendment (Social Media Minimum Age) Act 2024* (Cth), introducing a new Part 4A into the *Online Safety Act 2021* (Cth). This new Part requires service providers subject to this legislative obligation to take reasonable steps to prevent Australian children under 16 years of age from having accounts on their platforms from 10 December 2025.

Interaction with the Codes and Standards

While the social media age restrictions focus on preventing children under 16 years of age from having accounts on age-restricted social media platforms, the Codes and Standards primarily focus on preventing the generation, distribution, access or exposure to illegal material and children's access or exposure to age-inappropriate content.

Providers of age-restricted social media platforms will also have obligations under the Codes and Standards. Service providers that are not required to apply the social media age restrictions will still need comply with any relevant compliance measures under the Codes and Standards.

Service providers will need to consider whether their service is an age-restricted social media platform independently of considering which section of the online industry they fall under for the purposes of the Codes and Standards. Section 63C of the Act notes that an age-restricted social media platform may be, but is not necessarily, a social media service for the purposes of the Codes and Standards. This is because the definition of an age-restricted social media platform is broader than the definition of a social media service.[38]

This means that age-restricted social media platforms **can belong to different sections of industry** for the purposes of Codes and Standards. These are examples:

- An age-restricted social media platform may be a relevant electronic service for the purpose of the Codes and Standards.

[38] An age-restricted social media platform is defined as an electronic service that has **the sole purpose, or a significant** purpose, of enabling online social interaction between 2 or more end-users: section 63C(1)(a)(i) of the Act. A social media service is an electronic service that has the **sole or primary purpose** of enabling online social interaction between two or more end-users: section 13(1)(a)(i) of the Act.

- An age-restricted social media platform may be a relevant electronic service for the purpose of the Standards (Unlawful Material), but a social media service for the purpose of the Codes (Age-Restricted Material).

eSafety's Regulatory Guidance – Social Media Minimum Age contains additional information about how eSafety expects that industry may address their obligations under the Social Media Minimum Age obligation.

The Australian Government's approach to AI and keeping Australians safe

The protections provided under the Codes and Standards complement the Australian Government's broader focus on capturing the AI opportunity, spreading the benefits and keeping Australians safe, as outlined in Australia's National AI Plan released on 2 December 2025.[39]

The National AI Centre is supporting Australian industry with guidance and tools to adopt AI safety and secure productivity benefits. The Guidance for AI Adoption, released in October 2025, sets out six essential practices for responsible AI governance and adoption.[40] It offers practical and accessible steps to help organisations develop and deploy AI. There are two versions of the guidance targeting different AI maturity levels:

- Foundations: for organisations getting started in adopting AI
- Implementation practices: for governance professionals and technical experts.

Service providers should consider ways to implement the Guidance for AI Adoption in tandem with complying with the Codes and Standards.

The National AI Centre has also released guidance on Being Clear about AI-generated Content on 1 December 2025.[41] This guidance outlines the mechanisms that businesses can implement to be more transparent about their content, including labelling, watermarking and metadata recording.

The benefits of businesses being clear about content modified or generated by AI are that it can reduce regulatory risks, help to improve digital literacy across the community, and build trust and competitive advantage.

[39] Australian Government, Department of Industry, Science and Resources, National AI Plan (2 December 2025).
[40] National Artificial Intelligence Centre, Guidance for AI Adoption (21 October 2025).
[41] National Artificial Intelligence Centre, Being clear about AI-generated content (1 December 2025).

The Australian Government announced on 25 November 2025 that it will establish an Australian AI Safety Institute.[42] Its purpose will be to support Australia's regulatory approach to AI using existing, largely technology-neutral legal frameworks (including sector-specific guidance and standards) that can apply to AI and other emerging technologies,

The Institute will:

- monitor and test AI systems
- deliver clear and timely insights on emerging technology risks to government, industry, and the public
- work across government to ensure legislation and regulation keeps pace with technological advancements.

eSafety continues to work with the Department of Industry, Science and Resources to ensure alignment across our respective workstreams, which have different timescales and focus areas.

Guidance for providers of generative AI services or services with integrated AI features can be found at **Appendix D**.

[42] Australian Government, Department of Industry, Science and Resources, Australia establishes new institute to strengthen AI safety (25 November 2025).

Part 6: eSafety's approach to assessing compliance and deciding enforcement

eSafety continually monitors compliance with the Codes and Standards that are in force.

This monitoring will inform any decision eSafety makes to commence an investigation, where appropriate, and/or, in the case of codes, issue a direction to comply under the Act.[43]

eSafety may assess and investigate, on its own initiative or in response to complaints, whether a service provider has complied with the applicable codes or standards.[44]

eSafety can require the provision of relevant information, through examination or the production of documents from any person, for the purpose of an investigation under the Act.[45] A refusal or failure to provide the required documents may be subject to criminal or civil penalties where an appropriate exemption to the requirement cannot be demonstrated.[46]

Information eSafety will take into account

eSafety may take a range of information into account when monitoring and assessing the compliance of service providers with the Codes and Standards. These are some examples:

- Complaints made directly to eSafety about potential non-compliance with obligations.[47]
- Information from unresolved user complaints about potential non-compliance.
- Service providers' compliance reports provided to eSafety.
- Reports relating to technical feasibility and reasonable practicability of compliance relevant to certain provisions in the Standards (Unlawful Material) and Codes (Age-Restricted Material).
- Information obtained through eSafety's other regulatory mechanisms (such as the Basic Online Safety Expectations and complaints data about illegal and restricted online content).
- Information that service providers already publish voluntarily or as part of international transparency initiatives.

---

[43] Section 143 of the Act. A direction to comply does not apply to Standards.
[44] Sections 42(1)(f)-(g) of the Act
[45] See generally Part 14 of the Act.
[46] Section 205 of the Act.
[47] eSafety can receive complaints about potential code breaches under Section 40 of the Act.

- Information from stakeholders such as researchers, non-government organisations, law enforcement agencies and/or other governments, including international regulators.
- Information obtained through any routine assessment initiated by eSafety. For example, eSafety may check whether applicable services have appropriate age assurance or reporting and complaints mechanisms in place where required.
- Information obtained through desktop research, media reporting, social media posts, and other general information channels.

Complaints from the public

Most of the codes and standards contain provisions requiring service providers to be able to receive complaints from end-users about breaches of the service's obligations under the Codes and Standards.

eSafety can also receive complaints about non-compliance with a code or standard from Australian residents, including end-users of services and the general public. We also welcome complaints and engagement on compliance issues with researchers, civil society organisations and other expert groups. eSafety will use the information provided in complaints to identify and address potential systemic issues in the online industry, so we can help keep Australians safer online.

End-users in Australia seeking to make a complaint about non-compliance with a code or standard should be referred to eSafety's codes and standards complaint form.

What happens if a service provider is not complying with a code or standard?

eSafety takes a graduated approach, where appropriate, to compliance and enforcement. We strive to balance the protection of Australians against ensuring no undue burden is imposed on service providers.

In assessing a service provider's compliance with a code or standard, eSafety will consider whether the actions the service provider has taken fulfil the applicable requirements. Service providers are responsible for demonstrating that they meet the requirements of applicable obligations.

In some cases, eSafety may decide education and/or an informal request to seek rectification of a compliance issue is appropriate and likely to achieve compliance quickly.

In assessing compliance, eSafety:

- will take a fair and evidence-based approach
- will, across both phases of Codes and Standards, focus on the most serious harms and risks, and the most impactful requirements.
- will not assess compliance with optional requirements[48]
- may use information gathered from monitoring, assessment and enforcement action to identify specific priority areas for compliance during subsequent years
- may consider harms relevant to its strategic priorities for the relevant year, and will communicate any priority areas publicly to encourage proactive compliance.

eSafety will have regard to documented commitments by service providers to take specific measures to implement the Codes and Standards, where deficiencies may exist.

eSafety may approach a service provider to obtain further information and/or, where an investigation has commenced, use the information-gathering powers under the Act.[49] The steps taken by eSafety will depend on the nature of the potential breach being assessed and/or investigated, the information already available to eSafety and other factors and circumstances.

Step 1: Monitoring

eSafety actively monitors compliance with the Codes and Standards.

This can include:

- assessing complaints about potential non-compliance that eSafety has received
- compiling complaints about class 1 or class 2 material that have been received by eSafety
- engaging with experts, victims and law enforcement agencies
- directly assessing features and functions on a service.

Step 2: Investigation

Where appropriate, eSafety may conduct an investigation. The investigatory steps taken by eSafety will depend on the nature of the potential breach, the information already available to eSafety and others factors.

[48] Compliance with optional compliance measures may be taken into account by eSafety when considering whether or not a Code is deficient under Section 145 of the Act.
[49] Sections 199, 203 of the Act.

The investigation may include:

- approaching a service provider to voluntarily provide further information
- using information-gathering powers under the Act[50]
- continuing to obtain information from other sources (as outlined in Step 1).

Step 3: Determination

The Commissioner or relevant delegate determines whether they are satisfied that the service provider has contravened or is contravening a code or standard that applies to them.

Step 4: Enforcement

For contravention of a code or standard, the eSafety Commissioner or relevant delegate can use a range of enforcement actions available under the Act. These include:

- **Formal warning:** This can be issued to warn a service provider that they have failed to comply with the requirements of a code or standard.[51]
- **Written direction to comply:** A provider can be given a written direction to comply if eSafety is satisfied that it has contravened or is contravening the requirements of a code that applies to its service/s.[52] Failure to comply with a written direction may result in further action. Directions to comply do not apply to standards.
- **Enforceable undertaking:** A service provider may enter into an agreement with eSafety to ensure compliance with a code or standard. These undertakings can be enforced by a court.[53]
- **Injunction:** This is an order granted by the Federal Court of Australia or the Federal Circuit Court of Australia to compel a service provider to take certain actions, or to refrain from taking certain actions. An injunction is available where a service provider has not complied with a standard or a direction to comply with a code.[54]
- **Infringement notice:** This is a notice that sets out the particulars of an alleged contravention and specifies an amount to be paid. If it is not paid, eSafety may commence civil penalty proceedings. Infringement notices may be issued by eSafety and do not require the involvement of a court.[55]

---

[50] Sections 199, 203 of the Act.
[51] Sections 144, 147 of the Act.
[52] Section 143 of the Act.
[53] Section 164 of the Act.
[54] Section 165 of the Act.
[55] Subject to requirements in the *Regulatory Powers (Standard Provisions) Act 2014* (Cth).

- **Civil penalty order:** This is an order requiring payment of a financial penalty by a service provider that does not comply with a code or standard.[56] A civil penalty order can only be made by a court following civil penalty proceedings. It can be up to 30,000 penalty units for individuals and five times this for corporations.[57]
- **Seeking Federal Court orders to require a person to cease providing a social media service or internet carriage service:** eSafety may apply to the Federal Court of Australia to seek an order that a particular provider of a social media service, relevant electronic services or designated internet services stop providing that service in Australia, or for an internet service provider to stop supplying that service in Australia. eSafety will usually only pursue this option in relation to non-compliance with codes or standards where the continued operation of that social media service represents a significant community safety risk and where there is continuous and apparently wilful non-compliance.[58]

eSafety's Compliance and Enforcement Policy sets out more information regarding eSafety's approach and investigative powers.

Service providers may seek guidance and information from eSafety, noting the limitations around the advice eSafety can provide outlined in **Part 4**.

Step 5: Non-compliance with a direction to comply with a code, or non-compliance with a standard

If a service provider does not comply with a written direction or take action in response to a formal warning, eSafety will determine whether to take additional enforcement action.

Step 6: Enforcement action for non-compliance with a direction to comply with a code, or non-compliance with a standard

Enforcement action for non-compliance may result in a civil penalty of up to 30,000 penalty units for individuals and five times this for corporations.[59]

---

[56] Sections 162-163 of the Act.
[57] Section 143(2) of the Act. The monetary value of 1 penalty unit is $330 (at the date of this regulatory guidance). The maximum penalty ordered by a Court against a corporation (which can include online service providers) can be five times more than the maximum penalty ordered against an individual.
[58] To apply for the order, eSafety must be satisfied that a service failed to comply with a civil penalty provision under the Online Content Scheme (such as a written direction to comply with a Code) on two or more occasions over the past 12 months, and that continued operation of the service poses a significant community safety risk. To grant the order, the Federal Court of Australia must also be satisfied of those factors: Sections 156-159 of the Act.

Review rights for decisions relating to codes

A service provider may seek either internal review, or external review by the Administrative Review Tribunal, of certain actions taken by eSafety relating to codes.[60] The purpose of these review rights is to ensure that eSafety has made the correct and preferable decision on a case-by-case basis. The service provider named in the direction may seek review of the following decisions:

- Giving a direction to comply with a code
- Varying a direction to comply with a code
- Refusal to revoke a direction to comply with a code.[61]

An internal review may not always be appropriate, particularly if the direction has been given by the eSafety Commissioner. Additional information about seeking an internal review can be found on eSafety's website.

[60] Sections 220, 220A of the Act.
[61] Section 220(19) of the Act, referring to decisions under Section 143.

Appendix A: Pre-assessed and defined categories of Designated Internet Services and Relevant Electronic Services

How are services differentiated under the Designated Internet Services Standard (Unlawful Material)?

Pre-assessed categories

For pre-assessed designated internet service (**DIS**) categories, these risk profiles are ranked as either Tier 1 (highest risk) or Tier 3 (lowest risk). No services are pre-assessed as Tier 2 in the Designated Internet Services Standard (Unlawful Material). Pre-assessed categories in the Designated Internet Services Standard (Unlawful Material) are:

- High impact DIS (Tier 1)
- Classified DIS (Tier 3)
- General Purpose DIS (Tier 3)
- Enterprise DIS (Tier 3).

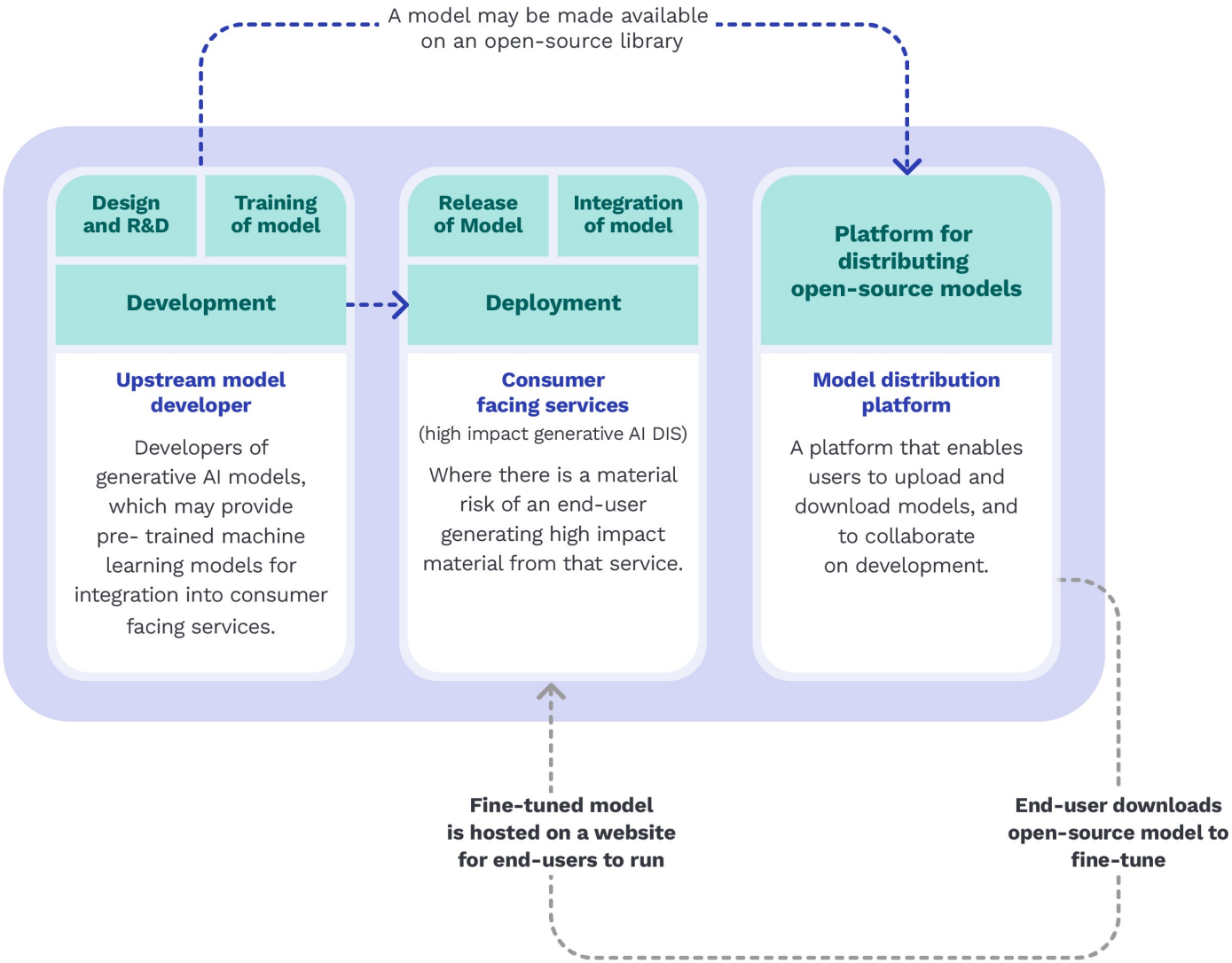
Defined categories

Noting that a broad range of websites and apps meet the 'Designated Internet Service' definition, the DIS Standard seeks to provide clarity by requiring specific measures for defined categories of services with unique risk profiles. As they have unique risk profiles, these defined categories are not attributed a risk tier. Defined categories in the DIS Standard are:

- End-user managed hosting service
- High impact generative AI DIS
- Model distribution platform.

The generative AI supply chain is shown in Figure 1. Only high-risk generative AI deployers and platforms distributing open-source models are assessed as defined categories. Upstream model developers are pre-assessed as Tier 3 in the Enterprise DIS category.

Figure 1: Treatment of generative AI services under the Designated Internet Services Standard (Unlawful Material)



How are services differentiated under the Designated Internet Services Code (Age-Restricted Material)?

Pre-assessed categories

For pre-assessed designated internet service categories, these risk profiles are ranked as either Tier 1 (highest risk) or Tier 3 (lowest risk). No services are pre-assessed as Tier 1 or Tier 2 in the Designated Internet Services Code (Age-Restricted Material).

Pre-assessed categories are in the Designated Internet Services Code (Age-Restricted Material) are:

- General Purpose Designated Internet Services (Tier 3)
- Enterprise Designated Internet Services (Tier 3).

No services are pre-assessed as Tier 1 or Tier 2.

Defined categories

Noting that a broad range of websites and apps meet the 'Designated Internet Service' definition, the Designated Internet Services Code (Age-Restricted Material) seeks to provide clarity by requiring specific measures for defined categories of services with unique risk profiles.

Unlike the Designated Internet Services Standard (Unlawful Material), certain services that fall within a defined categories in the Designated Internet Services Code (Age-Restricted Material) are still required to conduct a risk assessment, unless in lieu of performing a risk assessment, they have chosen to automatically assign themselves a Tier 1 risk profile. Defined categories in the Designated Internet Services Code that are required to conduct a risk assessment are:

- High impact generative AI Designated Internet Services
- High impact class 2 Designated Internet Services
- Model distribution platform.

Other defined categories in the Designated Internet Services Code not attributed to a risk tier are:

- End-user managed hosting service
- Model distribution platforms
- Classified Designated Internet Services.62

[62] A classified DIS can only be pre-assessed as tier 3 if it meets the requirements of Section 6(2) of the Designated Internet Services Standard (Unlawful Material).

How are services differentiated under the Relevant Electronic Services Standard?

Pre-assessed categories

Three categories in the Relevant Electronic Services (RES) Standard have been pre-assessed as being subject to the most comprehensive obligations. These are:

- Communication relevant electronic services
- Gaming service with communication functionality
- Dating services.

Defined categories

Three categories in the Relevant Electronic Services Standard have been defined with unique risk profiles. These are:

- Telephony Relevant Electronic Services
- Enterprise Relevant Electronic Services
- Gaming service with limited communications functionality.

Relevant Electronic Services that do not meet the definition of a pre-assessed or defined category must conduct a risk assessment to determine the service's requirements. Guidance on assessing risk is at **Appendix B**.

How are services differentiated under the Relevant Electronic Services Code (Age-Restricted Material)?

Pre-assessed categories

In the Relevant Electronic Services Code (Age-Restricted Material), the pre-assessed categories listed are subject to the most comprehensive obligations:

- Closed communication Relevant Electronic Services
- Other communication Relevant Electronic Services
- Gaming service with communication functionality
- Dating services.

Defined categories

In the Relevant Electronic Services Code (Age-Restricted Material), the defined categories of services listed have unique risk profiles:

- Telephony Relevant Electronic Services
- Enterprise Relevant Electronic Services
- Gaming service with limited communications functionality.

Appendix B: Risk profiles for Designated Internet Services and Relevant Electronic Services that are not pre-assessed or defined (Unlawful Material Standards)

Appendix B provides information in regard to the Designated Internet Services and Relevant Electronic Services Standards (Unlawful Material).

Risk assessment requirements for the Designated Internet Services and Relevant Electronic Services Codes (Age-Restricted Material) are located within the relevant Code.

Table 1: Designated internet services – risk profiles

Risk factor	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
Predominant purpose[63]	The purpose is provision of a general purpose DIS or a classified DIS.	The purpose is not to provide a general purpose DIS, a classified DIS or a high impact DIS.	The purpose is to enable end-users to post or access high impact materials.
Posting material[64]	The service: • does not enable end-users in Australia to post or view material to the service or a) enables end-users in Australia to post material only for the purposes of enabling such end-users to review or provide information on products, services, or physical points of interest or locations made available on the service	The service enables end-users in Australia to post and view material.	

[63] Designated Internet Services Standard (Unlawful Material) s 8(5)(a).
[64] Designated Internet Services Standard (Unlawful Material) s 8(5)(b).

Risk factor	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
	or b) enables end-users in Australia to post or share material only for the purpose of sharing that material with other end-users for a business, informational or government service or support purpose.		
Functionality – content creation[65]	The service: a) only makes available professionally produced material[66] to end-users and/or b) does not make available generative AI features.	The service makes available: • professionally produced material and end-user generated material and/or • generative AI functionality with a risk of producing material which would be classified as R18+ or lower.	The service predominantly makes available to Australian end-users: • material which has been posted by any end-user and/or • generative AI functionality with a risk of producing material which would be classified as R18+, X18+ or RC.[67]
Terms of arrangement for content acquisition[68]	The terms of arrangement (contracts) with third party providers of material to the service: a) prohibit class 1 material and b) require interventions to that ensure that ensure that the risk of class 1 material is immaterial. or c) due to the nature of the material provided under the terms of the arrangement by the	The terms of arrangement (contracts) with third party providers of material to the service do not: a) prohibit class 1 material and b) require robust methods to ensure that the risk of class 1 material is immaterial.	

[65] Designated Internet Services Standard (Unlawful Material) ss 8(5)(c), 8(5)(l)-(m).
[66] Professionally produced material is material produced by persons or entities who create such material:
- as a means of livelihood or for a commercial benefit; or
- on commission by the service provider (for example, a musical album by a professional musician, or a graphic design firm/photographer showcasing their portfolio).

[67] The requirements for a Tier 1 Designated Internet Service (DIS) and a high impact generative AI DIS are different. However, a service with the Tier 1 risk indicator of having functionality to produce R18+, X18 or RC material would likely also meet the definition of a high impact generative AI DIS.
[68] Designated Internet Services Standard (Unlawful Material) s 8(5)(g).

Risk factor	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
	third-party provider, there is an immaterial risk of class 1 material being provided.		
Visibility of material[69]	Any uploaded material is visible only to the Australian end-user and service provider.	Any uploaded material is available to the service provider and the user, and may be made visible and accessible to other end-users of the service.	
Terms of use[70]	The designated internet service has clear terms of use prohibiting the use of the service to solicit, access, generate, distribute or store (as applicable, having regard to the purpose and functionality of the service) class 1A material or class 1B material, and which give the designated internet service rights to enforce breaches of its terms of use.	The designated internet service has terms of use prohibiting the use of the service to solicit, access, generate, distribute or store (as applicable, having regard to the purpose and functionality of the service) class 1A material but not class 1B material. or The designated internet service does not have terms of use prohibiting the use of the service to solicit, access, generate, distribute and store (as applicable, having regarding the purpose and functionality of the service) either class 1A or 1B material	
Age of end-users[71]	The service is not likely to be accessed by children.	The service is likely to be accessed by children.	
Safety by design guidance and tools[72]	Other information from relevant Safety by Design assessments and guidance, should be used to inform appropriate risk tiering.		
Other factors	The list of factors that can be taken into account when carrying out a risk assessment under section 8 are non-exhaustive. Any other matter relevant to the service provider and the context in which they operate can form part of a risk assessment methodology.		

[69] Designated Internet Services Standard (Unlawful Material) s 8(5)(e).
[70] Designated Internet Services Standard (Unlawful Material) s 8(5)(f).
[71] Designated Internet Services Standard (Unlawful Material) s 8(5)(h).
[72] Designated Internet Services Standard (Unlawful Material) s 8(5)(j).

Table 2: Relevant electronic services – risk profiles

Risk factors	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
Purpose[73]	The service is predominantly for: • social interaction within a limited end-user group that has a pro-social common community interest (such as within a school, or neighbourhood or university community or a social or religious organisation or charity or sporting club or association) or • social interaction within a commercial or public enterprise that is limited to employees and/ or customers of the enterprise for the enterprise's stated purpose.	The predominant purpose of the service is to provide a forum for social interaction on a specific topic, such as to enable users to post reviews of products and services or for a limited commercial or public purpose. This may include the crowdfunding of commercial or charitable activities or social causes or to start an online petition for social change.	The predominant purpose is general social interaction, and it is not designed for social interaction in a specific context or for a specific purpose.
Function[74]	The service only enables: • sharing of material on a one-to-one basis between end-users, or within a defined group of end-users or • sharing of ephemeral material (material that lasts or is displayed only for a short time) without a sharing function. And the service does not have a chat or messaging service or a live streaming feature[75].		The service enables: • sharing and re-sharing of material to all end-users of the service and the material is permanent (not ephemeral) and/or • the service has a chat or messaging service, and/or enables live video streaming.

[73] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(a).
[74] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(b).
[75] Live streaming is live video that can be created and watched on a service by end-users in real time.

Risk factors	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
Terms of use[76]	The service has clear terms of use prohibiting the use of the service to solicit, access, generate, distribute or store (as applicable, having regard to the purpose and functionality of the service) class 1A material or class 1B material, and which give the service rights to enforce breaches of its terms of use.	The service has terms of use, which may not be clear, and/or do not prohibit the use of the service to solicit, access, generate, distribute or store (as applicable, having regard to the purpose and functionality of the service): a) class 1A material but not class 1B material or b) neither class 1A or class 1B material. and/or c) the service does not have terms of use which give the service rights to enforce breaches.	
Terms of arrangement for content acquisition[77]	The terms of arrangement (contracts) with third party providers of material to the service: a) prohibit class 1 material and b) require interventions to that ensure that ensure that the risk of class 1 material is immaterial or c) due to the nature of the material provided under the terms of the arrangement by the third-party provider, there is an immaterial risk of class 1 material being provided.	The terms of arrangement (contracts) with third party providers of material to the service do not: a) prohibit class 1 material and b) require robust methods to ensure that the risk of class 1 material is immaterial.	
Number of end-users in Australia that are monthly active end users[78]	1 to less than 500,000	500,000 to 3 million	Over 3 million
Likelihood of access by children[79]	The service is unlikely to be accessed by children.	The service is likely to be accessed by children.	

[76] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(c).
[77] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(d).
[78] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(b).
[79] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(e).

Risk factors	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
Safety by Design guidance and tools[80]	Other information from relevant Safety by Design assessments and guidance is used to inform appropriate risk tiering.		
Risk of generative AI material on the service[81]	The service does not make available generative AI functionality.	The service makes available generative AI functionality with a risk of producing material which would be classified as R18+ or lower.	The service makes available to Australian end-users generative AI functionality with a risk of producing material which would be classified as R18+, X18+ or RC.
Format of materials[82]	The service enables sharing of text or audio only.	The service: a) enables sharing of materials in text, image, audio and video and/or b) is enabled through immersive technologies.[83]	
Visibility of users[84]	The service **typically** only enables end-users to access and communicate with a list of contacts created by the end-user and does not enable end-users to: a) view or create a list of other end-users' individual connections on the service b) search for other end-users on the service using known identifiers (for example, name, username, email address) or connections c) search for other end-users on the service based on interests or keywords d) recommend other contacts to end-users based on interests or shared connections.	The service enables end-users to do any of the following: a) view or create a list of other users' individual connections on the service b) search for other end-users on the service using known identifiers, but not search for other end-users or discover material on the service based on interests or keywords c) recommend other contacts, or material, to end-users based on interests or shared connections.	The service enables end-users to: • search for and contact other end-users or discover material on the service based on interests or keywords or c) recommends other contacts or material to end-users based on interests or shared connections.

[80] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(g).
[81] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(h).
[82] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(a)-(b).
[83] Immersive technologies enable a user to experience and interact in three-dimensions (3D) with digital content in a way that looks, sounds and feels almost real. These technologies include augmented reality (AR), virtual reality (VR), mixed reality (MR) and haptics (interaction involving touch).
[84] Relevant Electronic Services Standard (Unlawful Material) s 8(5)(b).

Risk factors	Indicators of Tier 3	Indicators of Tier 2	Indicators of Tier 1
Other factors	The list of factors that can be taken into account when carrying out a risk assessment under section 8 are non-exhaustive. Any other matter relevant to the service provider and the context in which they operate can form part of a risk assessment methodology. For example, a relevant matter is also likely to be the number of global active end-users, as this heightens the risk that the service will be misused and the potential impact of dissemination of certain material, such as the child sexual abuse material of an Australian child.		

Appendix C: Summary of key risk assessments, communication and reporting requirements in the Codes and Standards

This table summarises the key risk assessment, communication and reporting requirements in the Codes and Standards outlined in **Parts 2 and 3**. Note that 'Phase 1' refers to the Unlawful Material Codes and 'Phase 2' refers to the Age-Restricted Material Codes.

	Designated Internet Services		Relevant Electronic Services		Social Media Services		Search Engine Services		App Distribution Services		Hosting Services		Internet Carriage Services		Equipment	
	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Conduct risk assessment	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓	✓
Provide risk profile information	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗	✓	✓
Update eSafety on relevant changes	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✓
Assess risk before making material changes	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓
Refer unresolved complaints to eSafety	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✓
Submit compliance reports	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Notify eSafety of app removal									✓	✗						

Appendix D: Guidance on terminology for the Codes and Standards

Systems, processes and technologies

The Codes and Standards refer to 'systems, processes and technologies' as the basis for some requirements. This reflects the intent of the codes and standards, which is to provide safeguards focused on minimising systemic online harms, rather than only the removal of individual items of material after they have already been shared.

The Codes and Standards contain requirements for service providers to implement appropriate systems, processes and technologies to, amongst other things:

- detect and remove known child sexual abuse and pro-terror material[85]
- detect and remove online pornography, self-harm material and high impact violence material[86]
- in some cases, prevent generative AI services from generating outputs that contain a generative AI restricted category of material.[87]

An exception applies if doing so would:

- not be technically feasible and reasonably practicable
- introduce a systemic weakness or vulnerability into the service
- (in relation to an end-to-end encrypted service) implement or build a new decryption capability into the service or render methods of encryption used in the service less effective.

Processes consist of a series of steps which should be documented and followed internally, and which set out how service providers may respond to risks, for example through standard operating procedures. Receiving and actioning user reports is an example of a process.

Technologies are capable of automatically taking certain actions, such as matching material against verified lists of child sexual abuse or pro-terror material.

---

[85] Designated Internet Services Standard (Unlawful Material) ss 20-21; Relevant Electronic Services Standard (Unlawful Material) ss 19–20.
[86] Social Media Services (Core Features) Code (Age-Restricted Material), compliance measures 8.1–8.3.
[87] Designated Internet Services Code (Age-Restricted Material), compliance measure 10.2; Social Media Services (Core Features) Code (Age-Restricted Material), compliance measure 10.2; Relevant Electronic Services Code (Age-Restricted Material), compliance measure 16.2.

Systems can encompass processes and technologies, as well as other inputs and outputs. For example, a system can integrate the use of hash matching technologies with processes for human review.

Technically feasible or reasonably practicable

In some cases, the Codes and Standards apply an exception if a requirement is not technically feasible or reasonably practicable. This includes obligations to implement:

- appropriate systems, processes and technologies, such as to detect and remove known child sexual abuse and pro-terror material, under the Unlawful Material Standards
- age assurance requirements under the Age-Restricted Material Codes.

Technically feasible

The term 'technically feasible' maintains its ordinary meaning. It is intended that service providers will first consider whether a system or technology is technically feasible before considering whether it is reasonably practicable.

The term 'technically feasible' does not contain any measure of practicality, proportionality or reasonableness. If it is possible under current technology for a person to do something, on this test taken in isolation, it will be technically feasible no matter the resources required to do so and any impacts on the person or business.

For providers of end-to-end encrypted service, please refer to the section on 'systemic weaknesses and vulnerabilities'.

Reasonably practicable

What is 'reasonably practicable' is determined objectively. This means that a service provider must meet the standard of behaviour expected of a reasonable person in the service provider's position who is required to comply with the same obligation.

In determining whether the system or technology is or is not reasonably practicable, any burden in addressing impediments to implementation must be balanced against the severity of risks and harms to end-users.

When determining if a measure is reasonably practicable, service providers should consider the risk of any child sexual abuse or pro-terror material being stored on, or distributed by or to, Australian end-users. Service providers should also consider whether the system or technology is proportionate to that risk, the costs and practicality of implementation and

whether the system or technology is likely to achieve the intended outcome of the Unlawful Material Standard or Age-Restricted Material Code.

Interaction between 'technically feasible' and 'reasonably practicable'

In assessing a system or technology, a service provider might find that its implementation is technically feasible, but that there are other significant impediments to implementation that do not justify its implementation in the circumstances where the risk of certain material is low. For example, it may be technically feasible for a start-up relevant electronic service or designated internet service provider to design their systems in a way which enables hash-matching to be deployed. However, it may not be reasonably practicable for an early-stage company to do so while their user base is still small and if the risk of harm is low.

Systemic weaknesses or vulnerability

Service providers are not required to implement systems or technologies to detect and remove material where doing so would require the service provider to implement or build a systemic weakness or systemic vulnerability into the service, or where it would require an end-to-end encrypted service to implement or build a new decryption capability or render methods of encryption used in the service less effective.

If a system or technology would make the encryption of an end-to-end encrypted service less effective, the Codes and Standards would not require it, and this applies whether a 'decryption capability' already exists or not.[88]

These exemptions in relation to systemic weaknesses and vulnerabilities are intended to complement those in the Online Safety (Basic Online Safety Expectations) Determination 2022 (Cth) (**BOSE Determination**).

[88] National Center for Missing and Exploited Children, Concentrix' Audit of NCMEC's Hash List (2024).

Defining systemic weakness and vulnerability

The terms 'weakness' and 'vulnerability' have significant cross-over. The Australian Cyber Security Centre defines a vulnerability as 'a weakness in a system's security requirements, design, implementation or operation that could be accidentally triggered or intentionally exploited and result in a violation of the system's security policy.'[89]

Many organisations will use industry standard definitions such as those from the National Institute for Standards and Technology, which defines a weakness as a 'defect or characteristic that may lead to undesirable behavior.'[90] It notes examples including missing a requirement or specification, having an architectural or design flaw, and an implementation weakness, including hardware or software defect.

The language of 'systemic' weakness and vulnerability recognises that there is no such thing as a perfectly secure service. Every service provider introduces theoretical weaknesses and vulnerabilities when they implement features for their own business purposes, but responsible service providers seek to build, design and test their services to minimise reasonably foreseeable risks. The explanatory statements recognise this, noting that risks must be 'actual and not merely theoretical'. 'Systemic' is intended to mean that the risk is a material weakness or vulnerability that effects the security of the system as a whole.

These provisions work in tandem with the protections for end-to-end encrypted services that also specify that service providers are not required to do anything that would result in a new decryption capability or render methods of encryption used in the service less effective. For the avoidance of doubt, this makes clear that service providers are not required to 'build back doors' or undermine end-to-end encryption.

Note: Due to the differences in purpose and technology between legislative schemes, the phrases 'systemic weakness' or 'systemic vulnerability' in the Standards (Unlawful Material) should not be interpreted using the definitions or caselaw relevant to Part 15 of the *Telecommunications Act 1997* (Cth). This is also emphasised in the explanatory statements to the Standards (Unlawful Material).[91]

---

[89] Australian Cyber Security Centre, Australian Signals Directorate, Vulnerability.
[90] Ross R, McEvilley and Winstead M, Engineering Trustworthy Secure Systems, National Institute of Standards and Technology (November 2022) p 63.
[91] Explanatory Statement of the Online Safety (Designated Internet Services—Class 1A and Class 1B Material) Industry Standard 2024; Explanatory Statement of the Online Safety (Relevant Electronic Services—Class 1A and Class 1B Material) Industry Standard 2024.

Appropriate alternative action

Certain service providers are required to implement appropriate systems, processes, and technologies to comply with obligations under the Codes and Standards. If it is not technically feasible or reasonably practicable for a service provider to do so, or doing so would build a systemic weakness/vulnerability or render encryption less effective, they must take appropriate alternative action.

When implementing an appropriate alternative action, for the avoidance of doubt, the service provider is not required to implement any action that would render encryption less effective, result in a systemic weakness or vulnerability, or require the implementation of a new decryption capability.

Several codes and standards refer to appropriate alternative action, as outlined:

Relevant Electronic Services Standard (Unlawful Material)

- Section 19: Detecting and removing known child sexual abuse material
- Section 20: Detecting and removing known pro-terror material

Designated Internet Services Standard (Unlawful Material)

- Section 20: Detecting and removing known child sexual abuse material
- Section 21: Detecting and removing known pro-terror material

App Distribution Services Code (Age-Restricted Material)

- Compliance measure 7.1: purchase/download restrictions for Australian children

The factors which must be considered when determining if something is appropriate are outlined in section 11 of both the Relevant Electronic Services and Designated Internet Services Standards. The appropriate alternative action may comprise a suite of additional steps, which when considered holistically in the context of the specific service, provide risk mitigations and appropriate safeguards in lieu of a technology or system. eSafety considers that these factors are also relevant when considering what constitutes appropriate alternative action under the App Distribution Services Code (Age-Restricted Material).

It is expected that service providers should assess the appropriate alternative actions that can be applied as part of a broader set of risk mitigations to protect the rights and best interests of children and end-users in Australia.

eSafety can request a report from a service provider describing where it is has deployed appropriate alternative actions with justification for the actions described.[92] It is therefore incumbent on the service provider to be capable of explaining how their alternative actions are appropriate in the circumstances.

Examples of appropriate alternative action for pro-terror material and child sexual exploitation material

eSafety cannot be prescriptive, as what is appropriate to do will depend on the service in question. However, service providers should use a combination of these measures and/or other suitable measures in a proportionate way. Some of these measures may also be steps that a service provider takes to disrupt and deter end-users from using the service for known and new pro-terror material and child sexual exploitation material.

- When product features create obstacles to reliably detect child sexual abuse and pro-terror material, carrying out and documenting risk assessments to ensure that risks are fully considered and safety measures are built into a service's design, rather than considered afterwards when harms arise.
- Providing features such as interstitial or warning messages and blurring potential child sexual abuse or pro-terror material.
- Providing easily accessible user reporting mechanisms (see also pages 43 to 46 for more details of requirements to enable reports and complaints). User reporting features that enable the provision of recent messages to the service provider. The number of messages shared with the service provider should be sufficient to enable the identification of the context of a communication, which may be particularly important for pro-terror material.
- Providing educational or supportive information, including safety information, to end-users.
- Using classifiers to detect signals and metadata relevant to unlawful and harmful material (such as behavioural signals related to private group membership, frequency of joining or leaving groups, engagement with children or young people using the service).
- For end-to-end encrypted services, using hashing, machine learning, artificial intelligence and other detection technologies on any parts of the service that are not end-to-end-encrypted (such as profile pictures, material in user reports, group names).

[92] See, e.g. Designated Internet Services Standard (Unlawful Material) s 32; Relevant Electronic Services Standard (Unlawful Material) s 33.

For end-to-end encrypted services, these measures align with section 8 of the BOSE Determination and the reasonable steps outlined in the Basic Online Safety Expectations Regulatory Guidance. These examples constitute important yet common measures which are already being deployed by some end-to-end encrypted services, as is shown in service provider responses to transparency notices.

Disrupting and deterring known and new child sexual exploitation and pro-terror material under Unlawful Material Standards

Certain providers of relevant electronic services and designated internet services must implement systems and processes and, if it is appropriate to do so, technologies that:[93]

- effectively disrupt attempts by end-users to use the service to solicit, generate, access, distribute or otherwise make available, or store, child sexual exploitation material or pro-terror material
- effectively deter end-users from using the service to solicit, generate, access, distribute or otherwise make available, or store, child sexual exploitation material or pro-terror material.

This requirement applies to both new material (which has not previously been verified) and known material (which has previously been verified). It ensures that service providers take meaningful steps to effectively disrupt and deter new and known child sexual exploitation and pro-terror material on their services, even if they are limited in their ability due to issues with technical feasibility, reasonable practicability, and systemic weakness/vulnerability.

The requirement is in addition to the complementary obligations to detect and remove known child sexual abuse and known pro-terror material. the 'technically feasible' and 'reasonably practicable' limitations which apply to the requirement to detect and remove known child sexual abuse and pro-terror material do not apply to the disrupt and deter requirement. It requires the service provider to implement systems and processes, but technologies need only be implemented **where it is appropriate to do so**.

This exception is in recognition that, at present, technologies to disrupt and deter new material may not be as accurate and robust as technologies to detect and remove known material, so service providers might encounter broader impediments in deploying such

[93] Designated Internet Services Standard (Unlawful Material) s 22; Relevant Electronic Services Standard (Unlawful Material) s 21.

technologies. Within this provision the use of 'appropriate' is included, which can include a consideration of proportionality.

Examples of disrupting and deterring known and new material

eSafety cannot be prescriptive, as there are a wide range of suitable systems, processes and technologies that vary depending on the service, and on the factors which must be considered when determining if something is appropriate as outlined in section 11 of the Unlawful Material Standards. These are some examples of steps that could be taken to disrupt and deter known and new material:

- The blocking of certain keywords and/or search terms that may be associated with child sexual exploitation or terrorism and violent extremism.
- Using machine learning to identify potential child sexual exploitation and terrorism and violent extremism.
- Using signals or indicators to prevent recidivism by end-users who have previously been banned or suspended for breaches of a service provider's terms of use for child sexual exploitation or terrorism and violent extremism.
- Providing end-users with clear communication advising if they are engaging in child sexual exploitation or pro-terror material, including conduct that violates terms of use (for example, providing a warning via a pop-up).
- Setting children's accounts to high privacy and safety settings by default to prevent offenders from contacting them.
- Deploying measures under the detect and remove requirements in relation to known material.[94]

As stated in the explanatory statements of the Unlawful Material Standards, in considering what an appropriate use of technology may be, service providers can consider their specific contexts and user base, including any underrepresented groups which may be at greater risk of technology systems falsely flagging their material. Service providers may also consider varying levels of accuracy which some machine learning classifiers have when classifying complex material at scale. As with any proactive technology, a system and process

[94] Measures to detect and remove known child sexual abuse and pro-terror material deployed under Section 20-21 of the Designated Internet Services Standard and Section 19-20 of the Relevant Electronic Services Standard can achieve part of the requirements of disrupting and deterring child sexual exploitation and pro-terror material. The requirements to disrupt and deter, however, is broader in scope as they apply to new and known child sexual exploitation material rather than only known child sexual abuse material.

incorporating human review of outputs of the technology, and the ability of end-users to appeal, are important in helping to mitigate limitations in a tool's accuracy and robustness.

Service providers should invest in and keep abreast of developments in the use of AI in trust and safety, to ensure that their measures are effective, safe and proportionate.

Continuous improvement

Under the Age-Restricted Material Codes, several service providers are required to regularly review their implementation of safety measures and continuously improve their implementation over time.

Designated Internet Services Code (Age-Restricted Material)

- Compliance measures 7.2 & 10.20: Continuous improvement of systems which can detect and action online pornography or self-harm material.
- Compliance measure 10.2: Continuous improvement of models with the aim of reducing unintentional use to generate a generative AI restricted category of material.

Relevant Electronic Services Code (Age-Restricted Material)

- Compliance measures 8.8, 9.8, 10.9, 11.8 & 15.7: Continuous improvement of safety features and settings.

Social Media Services (Core Features) Code (Age-Restricted Material)

- Compliance measures 8.1, 8.2 & 8.3: Continuous improvement of systems to detect and remove online pornography, self-harm material and high-impact violence material.

Social Media Services (Messaging Features) Code (Age-Restricted Material)

- Compliance measure 10.8: Continuous improvement of safety features and settings.

Internet Search Engine Services Code (Age-Restricted Material)

- Compliance measure 7.23: Continuous improvement of machine learning algorithms or models in reducing the risk of children accessing or being exposed to age-restricted material.

App Distribution Services Code (Age-Restricted Material)

- Compliance measure 7.6: Continuous improvement of safety tools and features.

To achieve continuous improvement, service providers should:

- use findings gathered from testing and monitoring systems to innovate and update their tooling
- invest in appropriate systems, tools and processes to support continuous improvement of these systems
- consider user feedback and reports about material and proactively update internal policies, tools and options to reflect feedback.

Reporting on compliance with continuous improvement obligations

As outlined in Part 3 of this regulatory guidance, service providers are also required to report to eSafety on their compliance with the Age-Restricted Material Codes either annually or upon request. This includes providing information about the steps they have taken to comply and why these steps were appropriate. When providing a compliance report relating to continuous improvement obligations, service providers should include relevant information and records about the following matters:

- Metrics outlining trends in the number of actions taken under compliance measures that must be continuously improved over time. This should include whether actions were automated or manually undertaken, and the number of complaints received in response to these actions.
- Where relevant, metrics relating to what material was reported to the service by end-users, and any trends over time. This should include:
    - how much material was reported
    - what form the material takes (such as images, videos and texts)
    - what category of age-restricted material the content falls under, and
    - what proportion of this material was actioned, whether this occurred manually or automatically, and how long it took to action reports.
- What steps service providers are undertaking to improve their compliance with continuous improvement obligations. This could include:
    - for generative AI services, any trends in the number of queries detected that request the generation of age-restricted material
    - records of consultations with crisis services, lived-experience groups and academic resources
    - records of any newly identified emerging trends relating to age-restricted material.

Appendix E: Guidance for providers of generative AI services

The Codes and Standards regulate class 1 and class 2 material, including when these materials are AI-generated.

AI-integrated Search Engine Services

Under the Search Engine Services Code (Unlawful Material) and Search Engine Services Code (Age-Restricted Material), service providers must consider the risks associated with any artificial intelligence features integrated into the search functionality[95] and take appropriate steps in relation these features.

Table 1: Applicable requirements for search engine services with integrated AI features

Search Engine Services Code	Requirements
Unlawful Material	• Reduce the accessibility and discoverability of class 1A material, including synthetic materials generated by artificial intelligence that may be accessible via the search engine service.[96] • Ensure that any AI features integrated into a search engine, such as longer form answers, summaries or materials, do not return search results that contain CSAM.[97] • Make clear when an end-user is interacting with any features using artificial intelligence, for example, to generate search results in the form of longer form answers, summaries or materials.[98]
Age-Restricted Material	• Provide information to end-users about how any internet search engine service features using generative artificial intelligence to generate longer form answers, summaries or materials, protects Australian children from exposure to online pornography and high-impact violence material.[99] • Improve systems, processes and/or technologies that aim to reduce the safety risks to end-users concerning synthetic materials generated by artificial intelligence that may be accessible via the internet search engine service.[100] Research detection technologies that assist end-users in identifying deep fake images that are accessible from the service.[101]

[95] Internet Search Engine Services Code (Unlawful Material), cl 5(c).
[96] Internet Search Engine Services Code (Unlawful Material), compliance measure 7(1).
[97] Internet Search Engine Services Code (Unlawful Material), compliance measure 7(2)(b).
[98] Internet Search Engine Services Code (Unlawful Material), compliance measure 7(9)(c).
[99] Internet Search Engine Services Code (-Restricted Material), compliance measure 7(19)(d).
[100] Internet Search Engine Services Code (-Restricted Material), compliance measure 7(23)(j).
[101] Internet Search Engine Services Code (-Restricted Material), compliance measure 7(23)(k).

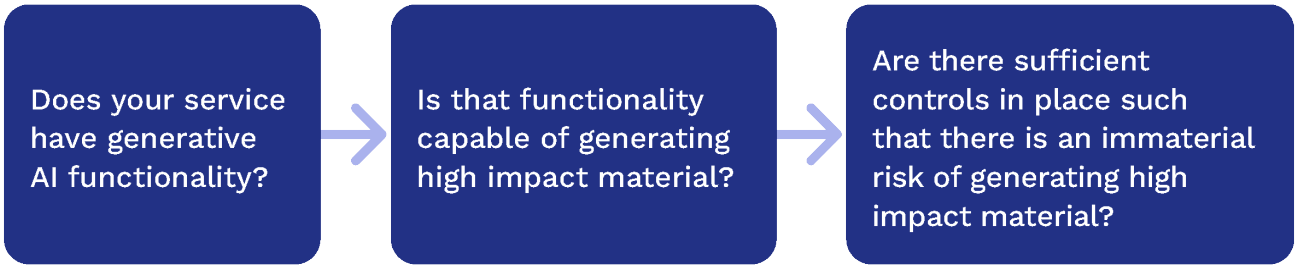
Unlawful Material Standards and Age-Restricted Material Codes: Requirements relating to 'high impact' generative AI services

Designated Internet Services (DIS)

A '**high impact generative AI DIS**' is defined as a designated internet service that uses machine learning models to enable an end-user to produce material and is capable of being used to generate relevant high impact material. What constitutes **relevant high impact material** differs depending on the applicable Code or Standard.

- **Unlawful Material: relevant high impact material** refers to material that would be classified under the National Classification Scheme as X18+ Restricted or RC,[102] or in the case of publication, classified as Category 2 Restricted or RC.[103]
- **Age-Restricted Material: generative AI restricted category of material** refers to online pornography, high impact sexually explicit material, self-harm-material, high impact violence material and violence instruction material.[104]

Figure 1: Questions to determine if a service is a high impact generative AI DIS



This category is intended to capture services which do not implement appropriate controls, safeguards or interventions to reduce the risk of the service being used to generate high impact material. If the risk of producing high impact material is 'immaterial', then the service provider will not be captured by the high impact generative AI DIS category.

[102] X18+ material consists of sexually explicit material that depicts actual (not simulated) sex between consenting adults, and which can be AI-generated. The threshold for a high impact generative AI DIS does not include R18+ material, for example high impact nudity.
[103] Designated Internet Services Standard (Unlawful Material) s 6(1).
[104] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 2.1.

Other services with integrated generative AI functionality

In addition to services dedicated to providing only a generative AI functionality to end-users, obligations also apply to services with an **integrated** generative AI functionality that can be used to produce high impact material. This includes completely new material, and material that has been created from editing existing material, such as some instances of 'deepfake' child sexual exploitation material.

For example, the Social Media Services (Core Features) and Relevant Electronic Services Codes (Age-Restricted Material) require providers of those services with integrated generative AI functionality to comply with the requirements in the respective Code, as a service with an AI companion chatbot feature. More information is provided below.

To determine whether a service is capable of being used to generate high impact material, service providers should assess the service as it is provided to end-users.

AI Companion Chatbot Feature – Age-Restricted Material Codes

The Social Media Services (Core Features) and Relevant Electronic Services Codes (Age-Restricted Material) create obligations for providers of services that have an AI companion chatbot feature.

An '**AI companion chatbot feature**' is defined as a feature on a social media service or relevant electronic service that:[105]

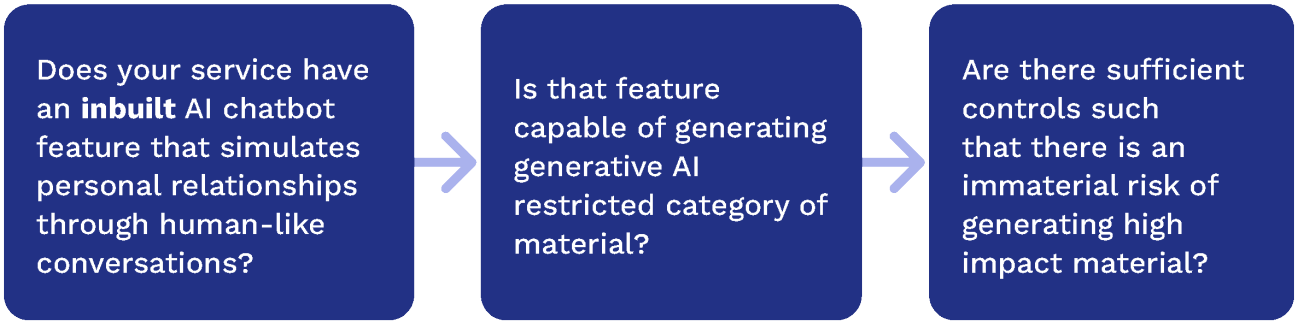
a) uses machine learning models to enable an end-user to produce material

b) is capable of being used to generate material in a generative AI restricted category and does not incorporate controls such that the risk of the feature being used to generate material in a generative AI restricted category is immaterial

c) is designed to simulate personal relationships through human-like conversations.

[105] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 2.1.

Figure 2: Questions to determine if a social media service or relevant electronic service includes an AI companion chatbot feature



This category is intended to capture services which do not implement appropriate controls, safeguards or interventions to reduce the risk of the service being used to generate high impact material. If the risk of producing high impact material is 'immaterial', then the service provider will not be captured by the AI companion chatbot feature definition.

Immaterial risk

A key element of the definition for a high impact generative AI DIS or AI Companion chatbot feature is whether the risk of generating synthetic high impact material using the service is 'immaterial'. Service providers should apply the following guidance on immaterial risk, to assess whether they are in scope,

An immaterial risk is one that is:

- likely to have insignificant consequences, or
- has a very low likelihood of occurring.

eSafety considers that there are almost always significant consequences that flow from the risk of a generative AI service being used to generate high impact material that is from and is class 1 and class 2 material. Therefore, the **likelihood** of a service being used in this way is the most relevant aspect for service providers to consider when assessing whether a risk is immaterial.

This means that the risk that end-users can generate synthetic high impact material is only 'immaterial' in relation to a service if there is a very low likelihood of this occurring on the service.

Service providers may deploy a range of design features and controls which lead to an immaterial risk of generating high impact material. eSafety's Generative AI Tech Trends Position Statement (2025) details the risk mitigations that can be deployed across the full generative AI lifecycle. The report on Safety by Design for Generative AI: Preventing Child

Sexual Abuse (2024), developed by Thorn and All Tech is Human, also set out relevant risk mitigations.

eSafety considers that for a general-purpose model to effectively mitigate the risk of generating high impact material, among other measures, relevant matching tools and classifiers should be deployed on training data, in user prompts and model outputs. The most effective solutions will apply safety interventions across each stage – in training data, user prompts and outputs.

Types of services considered as a high impact generative AI DIS or a companion chatbot AI feature

The risks arising from generative AI can be categorised in three ways:

- Inadvertently without explicit prompting – this occurs when a system unintentionally causes harm by generating incorrect or harmful responses.
- When explicitly prompted to do so – this happens when a model can be exploited for harmful purposes.
- The system has been optimised specifically for the purpose of generating high impact material.

Generative AI services which have been optimised or fine-tuned for high impact material – such as bespoke pornography generators and 'nudify' type models – present a higher risk of generating high impact material. Additionally, services that are interactive in nature, such as virtual girlfriend apps, are more likely to be high-risk. Interactivity is recognised as a factor that may increase impact under the National Classification Scheme.[106] Other services without clear terms of service and processes limiting the generation of potentially high impact materials are also likely to present a material risk.

Even general-purpose services which clearly prohibit the generation of high impact materials typically involve some risk of perpetrators circumventing safeguards to use a model to produce class 1 and class 2 material.

[106] Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, Australian Classification, How a rating is decided.

Generative AI services outside the scope of the high impact generative AI DIS category

Any provider of a DIS with generative AI features that does not fall within the definition of a high impact generative AI DIS or other defined category is still required to conduct a risk assessment under:

- subsection 7(1) of the Designated Internet Services Standard (Unlawful Material)
- clause 4.2 of the Designated Internet Services Code (Age-Restricted Material).

Providers of these services are still required to comply with obligations under the Codes and Standards dependent on their assessed risk profile.

Appendix F: Appropriate age assurance under the Age-Restricted Material Codes

Introduction

Overview

Under the Age-Restricted Material Codes, providers of certain high-risk services will be required to implement appropriate age assurance measures and then take steps to prevent children from accessing or being exposed to class 1C and class 2 material on their service.

This appendix outlines what constitutes appropriate age assurance under the Age-Restricted Material Codes, including:

- appropriate age assurance guidance
- steps to prevent children from accessing or being exposed to class 1C and class 2 material
- reporting on compliance with age assurance measures
- summary of age assurance requirements under the Age-Restricted Material Codes.

Age assurance in the Age-Restricted Material Codes and other regulatory schemes

Providers of certain online services are required to implement age assurance across several regulatory schemes in Australia, including the Social Media Minimum Age obligation, the Basic Online Safety Expectations (**Expectations**) and the Restricted Access System Declaration.

Social Media Minimum Age obligation

The *Online Safety Amendment (Social Media Minimum Age) Act 2024* was passed by the Australian Parliament in November 2024. From 10 December 2025 'age-restricted social media platforms' are required to take reasonable steps to prevent Australians under 16 from having accounts. The Age-Restricted Material Codes were drafted by industry associations in Australia in 2024-2025 to be applicable to multiple relevant sections of industry identified in the Act.

To support regulatory coherence, this regulatory guidance is closely aligned with the Social Media Minimum Age Regulatory Guidance as principles-based guidance, which was widely

consulted upon. However, there are a number of differences between the requirements under the Age-Restricted Material Codes and the Social Media Age Restrictions obligation, which are noted in this Regulatory Guidance where relevant – **service providers should ensure they follow the correct guidance for each.**

Restricted Access System Declaration

As outlined in Part 3, social media services, relevant electronic services and designated internet services provided from Australia are required to implement access-control systems under the Restricted Access System Declaration to restrict children's access to R18+ material online, upon receiving a notice from eSafety.

Table 1 provides an overview of the distinctions between the Age-Restricted Material Codes and other regulatory schemes.

Table 1: Distinctions between the Age-Restricted Material Codes and other regulatory schemes

Feature of scheme	Age-Restricted Material Codes	Social Media Minimum Age obligation	Restricted Access System
What services are required to conduct age assurance?[107]	High risk services across eight sections of the online industry, as outlined in Table A	Age-restricted social media platforms	Designated internet services, relevant electronic services and social media services provided from Australia
What is the relevant age threshold?	Children in Australia: aged under 18	Age-restricted users: aged under 16	Children in Australia: aged under 18
Explanation of users	End-users in Australia	End-users who have **accounts** who are ordinarily resident in Australia	Applicant: a person who makes a request for access to relevant class 2 material
What is the threshold for age assurance measures?	Appropriate age assurance	Reasonable steps to prevent age-restricted users from having accounts	Reasonable steps to confirm an applicant is at least 18 years of age.

[107] Some services may be required to implement age assurance under both the Codes (-Restricted Material) and the Social Media Minimum Age. Section 5.1(c)(iv) of the Head Terms to the -Restricted Material Codes requires service providers to consider the interaction between measures which require age assurance in the relevant Code, and other applicable Australian laws which may achieve the online safety objectives while minimising the collection of personal information.

What is required when a user is found to be under the relevant age?	Implementation of access control measures preventing children from accessing high-risk material implemented in line with requirements under the relevant Code, as outlined in Table A	• Deactivation or removal of accounts • Prevention of age-restricted users from creating new accounts	Implementation of access control measures preventing children from accessing class 2 material
Privacy obligations	Compliance with the *Privacy Act 1988* (Cth) and the Australian Privacy Principles	Compliance with the *Privacy Act 1988* (Cth), the Australian Privacy Principles and s 63F of the *Online Safety Act*.	Compliance with the *Privacy Act 1988* (Cth) and the Australian Privacy Principles

Basic Online Safety Expectations

As outlined in Part 3, the Expectations set out the steps the Australian Government expects should be taken by providers of designated internet services, relevant electronic services and social media services to keep end-users in Australia safe online. **The Expectations are not enforceable, and so are different from the other schemes outlined above. However, compliance with the Expectations is related to a number of relevant information gathering and publishing powers under the Act which are enforceable.**

This includes section 12 of the Determination which details the expectation that service providers will take reasonable steps to ensure that technological or other measures are in effect to prevent access by children to class 2 material provided on the service. An example of a reasonable step could be to implement appropriate age assurance measures. For further information, please see the regulatory guidance to the Expectations.

Appropriate age assurance guidance

Online safety objectives of the Age-Restricted Material Codes

The Age-Restricted Material Codes have two online safety objectives:[108]

a) **Objective 1:** Protect and prevent children in Australia from accessing or being exposed to class 1C and class 2 material[109]

b) **Objective 2:** Provide Australian end-users with effective information, tools and options to limit access and exposure to class 1C and class 2 material.

[108] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 4.
[109] eSafety, Online Content Scheme Regulatory Guidance (January 2025) p 4.

Age assurance requirements in the codes reflect the risk that a provider's service could be used to expose children to class 1C and class 2 material on that service. These requirements, detailed in this appendix, are **technology neutral**, allowing service providers to choose how best to meet the required outcomes within their existing framework of operations.

A summary of age assurance requirements under the Age-Restricted Material Codes can be found at **Table A of this Appendix**.

Appropriate age assurance

For the purposes of the Age-Restricted Material Codes, high-risk services are required to implement appropriate age assurance measures to assess whether an Australian end-user is **at least 18 years of age**. They are then required to implement **access control measures** which prevents these end users from encountering class 1C and class 2 material on these services.

Australian end-users

Under the Head Terms to the Age-Restricted Material Codes, providers of services are required to take steps to address the presence of this age-inappropriate online content on their services for all people in Australia who are users of the service.[110]

Under the Social Media Minimum Age obligation, service providers are required to consider whether end-users are ordinarily resident in Australia. Further information on this can be found on our website.

Clause 5.1(c) of the Head Terms to the Age-Restricted Material Codes outlines requirements for what constitutes **appropriate age assurance**. This includes that service providers should consider:

1. the technical accuracy, robustness, reliability and fairness of the solution[111]
2. whether the age assurance measure/s have been designed to comply with Australian privacy laws[112]

[110] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 4. This is the scope adopted by industry associations drafting the -Restricted Material Codes.
[111] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(i).
[112] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(iii).

3. whether the impact on user privacy of any such measures for a service is proportionate to the online safety objectives of Age-Restricted Material Codes [113]
4. how to minimise the collection of personal information if a service provider is:
    - subject to both the Age-Restricted Material Codes and other applicable Australian laws which may require age assurance for the same product or service[114]
    - subject to age assurance requirements under the Age-Restricted Material Codes for multiple products or services that they provide.[115]

Examples of appropriate age assurance

Clause 5.1(c) of the Head Terms to the Age-Restricted Material Codes provides a non-exhaustive list of examples of forms of appropriate age assurance for the purposes of the Age-Restricted Material Codes. This includes:

a) matching of photo identification

b) facial age estimation

c) credit card checks

d) digital identity wallets or systems

e) attestation by a parent of age or whether an Australian end-user is a child

f) use of artificial intelligence technology to estimate age based on relevant data inputs;

g) other measures meeting the requirements of section 8 (Confirmation of age) of the *Online Safety (Restricted Access Systems) Declaration 2022*

h) relying upon appropriate age assurance measures implemented in respect of the relevant end-user by:
    - another party (whether another industry participant, government agency, a third-party vendor or another third party) and confirmed by an age signal or other mechanism provided to the service provider by that other party
    - the service provider in respect of another service as contemplated in clause 5.1(c)(vi), although this list is not comprehensive and additional age assurance measures may also be considered appropriate.

---

[113] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(iii).
[114] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(iv).
[115] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(iv).

eSafety does not recommend specific providers of age assurance measures, or mandate the use of any particular type of age assurance technology. However, service providers are encouraged to consider:

- methods of age assurance that have been independently certified or accredited against relevant international and domestic standards on matters such as accuracy, security and fraud resilience[116]
- the findings of the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts (DITRDCSA) Age Assurance Technology Trial.

Examples of age assurance measures that are not considered appropriate age assurance

Section 5.1(c) of the Head Terms to the Age-Restricted Material Codes lists examples of age assurance measures that will **not** be considered appropriate for the purposes of the Age-Restricted Material Codes. This includes:

a) requiring a user to self-declare their own age or whether they are a child (without more)

b) contractual restrictions on the use of the relevant service by children (without additional measures).

Technically accurate, robust, reliable and fair

Service providers should ensure age assurance measures – and the systems and processes surrounding them – are **technically accurate, robust, reliable and fair**.

Technically accurate and reliable

Age assurance systems should reliably produce a result that provides a service provider with a sufficient level of confidence as to whether an end-user is at least 18 years of age.

To enable a flexible and proportionate approach, **eSafety does not propose a minimum accuracy level** for age assurance methods. Services providers should determine if the measures or combination of measures implemented gives them sufficient confidence of an end-users age:

- Service providers should **define acceptable error thresholds based on their risk, service type, and user base**. Service providers are encouraged to consider relevant international standards and accreditation schemes to inform their consideration of accuracy levels

[116] For example, ISO/IEC 27001 Information security, cybersecurity and privacy protection – Information security management systems - Requirements and ISO/IEC 30107 Information technology – Biometric presentation attack detection Part 1: Framework.

- Service providers are not required to eliminate all uncertainty but should seek to minimise harm and ensure decisions are proportionate, fair, and reviewable. This includes the ability to report on error rates and work to continuously improve age assurance methods.

Where age assurance is based on inference or estimation and a **buffer threshold** is set, service providers should ensure that the threshold is appropriately configured, having considered the accuracy of the underlying technology, the confidence of the estimation or inference, and the risk of unreasonably over-blocking users that are not age-restricted users.

All age assurance methods should be backed by accessible, timely and accurate appeal and review processes.

Robustness

Service providers should implement age assurance systems that are secure and reasonably resistant to circumvention, and ensure their own systems and processes are also sufficiently robust to withstand such challenges.

- Service providers should ensure any age assurance method employed has undergone sufficient testing and evaluation before use and while in use. Service providers are encouraged to **undertake and document ongoing internal testing procedures as well as seek external audits or independently validated testing** to support transparency.

Examples of measures service providers can take include:

- conducting periodic **red-team testing**,[117] including simulating bypass attempts by underage end-users
- procuring **third-party audits**[118] of age assurance and complementary measures, including circumvention controls.

Service providers should ensure that review and evaluation are conducted regularly and in response to any material changes on the service. Evaluation criteria, outcomes, and processes should be recorded to demonstrate they have implemented appropriate age assurance and be provided to eSafety in line with reporting requirements.

Further information about requirements to prevent circumvention can be found in the section, Reasonable steps to prevent circumvention.

---

[117] Testing to see if something can be circumvented, bypassed or tricked. A red team is a group of people authorised and organised to emulate a potential adversary's attack or exploitation capabilities against an enterprise's security posture. National Institute of Standards and Technology (NIST), *Information Technology Laboratory Computer Security Resource Center – Glossary*, (2025).

[118] An assessment conducted by an independent, external entity.

Fair

Service providers should ensure age assurance methods and surrounding systems and processes are fair. This includes the requirement that age assurance methods and systems should be accessible and inclusive.

To achieve this, eSafety expects service providers to consider the range of existing and prospective Australian end-users with **diversity in appearance, abilities and capacities**, and implement systems and safeguards to ensure their methods are accessible and produce outcomes that are inclusive and fair for **all end-users**.

The use of age assurance should not unfairly inhibit access for certain end-users or impact certain groups disproportionately without adequate mitigations and support to minimise the potential for bias and discrimination.

eSafety expects service providers to take a number of steps to ensure that age assurance methods and surrounding systems and processes are fair:

- Service providers should test their age assurance methods in the Australian context, including by looking at **different demographics within Australia** and whether the age assurance system is accessible, inclusive and fair for these demographics. Accuracy should be evaluated and recorded across different cohorts, with an aim to minimise bias and improve consistency in results over time.
- Service providers should **mitigate the impact of accessibility or bias issues** in the age assurance methods they use and **build processes to support those who may be adversely affected**. This includes ensuring that systems are inclusive of the diverse needs of communities across Australia — such as Aboriginal and Torres Strait Islander peoples, culturally and linguistically diverse communities, and those with limited access to digital infrastructure or identity documentation. This is particularly relevant where age assurance methods are based on machine learning or involve automated decision making.
- Service providers should produce **clear and easy-to-understand information** about their age assurance methods. This information should be made available for users at a range of literacy levels and in a variety of different languages. In developing this information, providers should also ensure alignment with Web Content Accessibility Guidelines (WCAG) 2.1.
- Service providers should provide an accessible method for end-users to request a review of any age assurance decisions. Age assurance methods should be adaptable to individual end-user circumstances, needs and preferences.
- Service providers should offer a choice between a range of age assurance methods, giving end-users flexibility and agency in choosing methods that best suit their circumstances.

- Service providers should account for those who do not have access to documents, are facing challenging circumstances or experiencing vulnerability, or otherwise face barriers engaging with age assurance methods. This should include accepting a range of options rather than a narrow list of age documents and providing non-document-based options.
- Where appropriate, service providers may consider methods such as professional or community vouching, or assessment of alternative evidence of age, where end-users have been unable or unwilling to use other provided methods of age assurance. Service providers should consider whether this is **reasonable** in the circumstances and ensure such methods are supported by appropriate validity checks.

Use of government ID as a sole method of age assurance

The Online Safety Act sets out that under the Social Media Minimum Age obligation, age-restricted social media platforms are not permitted to use government identification documents as the sole method of age assurance.[119]

While this requirement is not present under the Age-Restricted Material Codes, eSafety considers that an age assurance mechanism where the use of government identification documents is the sole method to verify a user's age will likely not satisfy the principle in 5.1(c)(i) of the Head Terms to the Age-Restricted Material Codes that services providers should consider fairness in applying appropriate age assurance, having regard to the matters detailed above.

Technically feasible or reasonably practicable exception

Service providers that are required to implement age assurance measures must do so where it is **technically feasible** and **reasonably practicable**.

For more details about the concepts of technical feasibility and reasonable practicability in the Codes and Standards, see **Appendix D**.

Privacy and age assurance under the Age-Restricted Material Codes

Privacy and the protection of personal information are important for everyone's agency, dignity, and safety.[120] eSafety considers that privacy and online safety are not mutually

[119] Section 63DB of the Act.
[120] Article 17 of the *International Covenant on Civil and Political Rights* (ICCPR) provides that no one shall be subjected to arbitrary or unlawful interference with his or her privacy, family, home or correspondence, nor to

exclusive objectives and age assurance measures can be implemented in ways that protect user privacy.

Clause 5.1(c)(iii) of the Head Terms to the Age-Restricted Material Codes notes that when implementing appropriate age assurance measures, service providers must:

1. consider whether their age assurance measures **comply with Australian Privacy Law,** including the Privacy Act and the Australian Privacy Principles regulated by the Office of the Australian Information Commissioner (**OAIC**)
2. consider whether the impact on user privacy of age assurance measures is **proportionate** to the online safety objectives of the Age-Restricted Material Codes.

Clauses 5.1(c)(iv) and (vi) of the Head Terms to the Age-Restricted Material Codes also provide service providers with obligations to **minimise the collection of end-user personal data**.

Clause 6 of the Head Terms to the Age-Restricted Material Codes also say that the Codes (and accordingly, the application of any form of relevant appropriate age assurance measures) **do not** require service providers to do any of the following:

1. Implement or build a systemic weakness into a form of encrypted service or other information security measure.
2. Undertake monitoring of private communications between end-users.
3. Verify or publish the real identity of any end-user.
4. use or disclose personal information of an Australian end-user (or do anything else) that would put the industry participant in breach of an applicable privacy law.
5. Use or disclose personal information of a foreign end-user in a way that would put the industry participant in breach of any foreign law relating to the management of personal information of that end-user.
6. Take any other action within Australia that is prohibited under Australian law or regulation by which the industry participant is bound.

Compliance with Australian Privacy Law

Service providers have a number of obligations under Australian Privacy Law, including **the Privacy Act** and the **Australian Privacy Principles**.

---

unlawful attacks on his or her honour or reputation, and that everyone has the right to the protection of the law against such interference or attacks. For interference with privacy not to be arbitrary, it must be lawful and in accordance with the provisions, aims and objective of the ICCPR and should be reasonable in the particular circumstances.

Service providers could choose to conduct a privacy impact assessment of any age assurance measures they implement, which will assist with the service provider's assessment of both positive and negative privacy impacts of those measures, and with ensuring that relevant collection notices and other steps are taken, as required by Privacy Laws. This may inform their approach to 'appropriate age assurance' as contemplated by Clause 5 of the Head Terms.

Service providers should aim to gather the minimum information and data needed to make decisions appropriate for their service and circumstances. Policies should be calibrated to ensure the collection, use and retention of personal information is reasonably necessary and proportionate. Service providers are strongly encouraged to use non-personal information as far as possible, and avoid handling of **sensitive personal information**.[121]

eSafety **does not expect service providers to retain personal information as a record of individual age checks** for the purpose of reporting on compliance with the Age-Restricted Material Codes.

The Privacy Act and the Australian Privacy Principles are regulated by the Office of the Australian Information Commissioner (**OAIC**). Service providers should have regard to any privacy and related guidance released by the OAIC. If service providers have queries regarding their requirements under these privacy regulations, they should contact the OAIC.

The Privacy Act, including the Australian Privacy Principles and any industry codes that may be made in accordance with the Privacy Act, are the applicable Australian Privacy Laws for the purpose of the Age-Restricted Material Codes.

The OAIC has published guidance on a service providers' privacy obligations under the Australian Privacy Principles in relation to the implementation of age assurance systems. Service providers should refer to this guidance for further details about their privacy obligations.

eSafety notes that that the Social Media Minimum Age obligation introduced additional privacy-related provisions for age-restricted social media platforms located within the *Online Safety Act 2021*. Service providers should refer to eSafety's Social Media Minimum Age Regulatory Guidance the OAIC's Privacy Guidance for age-restricted social media services for further details.

[121] Office of the Australian Information Commissioner, Australian Privacy Principles Guidelines, Chapter B: Key concepts (21 December 2022).

Consideration of whether the impact on user privacy is proportionate to the online safety objectives of the Age-Restricted Material Codes

As required under clause 5.1(c)(iii) of the Head Terms to the Age-Restricted Material Codes, service providers must also consider whether the impact on user privacy of age assurance measures is proportionate to the **online safety objectives of the Age-Restricted Material Codes.**

This includes considering the **balance** of the measures they implement having regard to the risk of the harm they mitigate and the impact they have on end-users.

Risk of harm

The Age-Restricted Material Codes require service providers to implement risk-based obligations. Clause 1.1(c) of the Head Terms to the Age-Restricted Material Codes provides that compliance measures adopted under the Age-Restricted Material Codes should be proportionate to the level of risk associated with the online service and to the size and capacity of the industry participant responsible for the online service.

The Age-Restricted Material Codes generally require providers of services with the highest risk of enabling children to access or be exposed to class 1C and class 2 material to implement appropriate age assurance measures. eSafety considers that service providers are able to use risk assessments to identify the risk of harm that age assurance measures would mitigate if implemented.

There are two types of risk assessments in the Age-Restricted Material Codes:

1. Certain service providers are required to assess the risk that their service will be used to enable children to access or be exposed to class 1C and class 2 material or otherwise nominate themselves as high-risk.
2. Services that fall within this high-risk category can assess that the risk of harm that age assurance measures would mitigate is **high**.

Certain service providers must conduct a risk assessment and take their service's risk level into account when developing appropriate compliance measures under the Age-Restricted Material Codes. For example, providers of search engine services with a higher risk profile may be required to implement more stringent age assurance measures. Accordingly, if a provider of a search engine service with a higher risk profile uses age inference methods, they should have reduced buffer thresholds in comparison to lower risk search engine services.[122]

[122] Internet Search Engine Services Code (-Restricted Material) cl 5.

Impact on users

After assessing the risk of harm that the age assurance measures would mitigate if implemented, service providers must consider whether the risk mitigated counterbalances the impact of any age assurance measures on users.

All service providers should avoid **unreasonable practices that risk over-blocking access or infringing on the rights of Australians**. There are several considerations for service providers:

- It may be more appropriate for providers of higher-risk services, such as those with the sole or predominant purpose of enabling access to class 1C and class 2 material, to conduct appropriate age assurance and implement associated access control measures for *all users* of their services so that they have the highest degree of certainty possible about the age of users.
- If a service provider can use existing user data (in circumstances where that data has been collected such that it can be legally applied to that purpose) to infer with reasonable confidence that certain end-users are over 18,[123] then it may be reasonable and appropriate not to require those users to conduct appropriate age assurance again.
- If a service provider provides multiple services which are required to implement appropriate age assurance measures, then it may be more appropriate for that service provider to share an age signal or other age-related settings (see clause 5.1(c)(vi) of the Head Terms) rather than seeking that end-users complete appropriate age assurance multiple times.

Service providers should also consider if age assurance measures disproportionately impact certain members of the Australian community. This could be done, for example, by appropriately engaging with safety and community organisations to inform them about the effectiveness of their age assurance measures.

Minimising end-user data collection where multiple requirements apply

The Age-Restricted Material Codes operate in parallel with other regulations that require the assessment of a user's age online. Accordingly, service providers may be required to comply with multiple obligations to age assure users. eSafety considers that service providers should, where possible, minimise end user data collection if this is the case.

Where other applicable Australian laws require age assurance for the same product or service

[123] For example, the length of time an account has been held. See also Age Check Certification Scheme and Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, *Age Assurance Technology Trial – Report* (2025), Part E – Age Inference

Clause 5.1(c)(iv) of the Head Terms to the Age-Restricted Material Codes requires service providers who must comply with multiple obligations to age assure users for the same service, to consider how to best achieve the online safety objectives of the Age-Restricted Material Codes **whilst minimising the collection of personal information**.

Service providers required to conduct age assurance under both the Social Media Minimum Age obligation and the Age-Restricted Material Codes

eSafety notes some service providers may be required to conduct age assurance in accordance with both the Social Media Minimum Age obligation and the Age-Restricted Material Codes. In accordance with the Head Terms, these services must consider how they intended to comply with any parallel obligations under Section 63D of the *Online Safety Act 2021* (Cth) and any relevant age assurance measures in an applicable Age-Restricted Material Code while minimising the collection of personal information.

Where service providers have multiple services subject to age assurance requirements under the Age-Restricted Material Codes

Service providers with multiple services subject to age assurance measures in the Age-Restricted Material Codes should consider how to reduce the impact of these measures on their users, including by minimising the collection of personal information.

Clause 5.1(c)(vi) of the Head Terms provides the option to these service providers of sharing an age signal or other age-related settings generated by the age assurance measures in place for the other service, whether via a centralised account or otherwise. Service providers are required to ensure that the relevant end-user has agreed for these signals or settings to be shared between services if undertaking this option.

Where possible, eSafety considers service providers should also consider whether they are able to accept third-party age signals if those age signals are also based off a method of appropriate age assurance as contemplated in clause 5.1(c)(vi)(H) of the Head Terms.[124]

[124] For example, sharing an age range: Apple, Declared Age Range (2025).

Improving age assurance measures over time

Service providers have obligations to improve age assurance measures over time.

- All service providers are required to comply with clause 1.3 of the Head Terms to the Age-Restricted Material Codes regarding ongoing work on age assurance.
- Certain service providers have test, monitor, and (in some cases) improve their age assurance measures over time.

Ongoing work on age assurance measures

As outlined in clause 1.3 of the Head Terms to the Age-Restricted Material Codes, it is expected that all industry participants will continue to look for ways to collaborate and contribute proactively to implement appropriate controls to protect children from accessing or being exposed to class 1C and class 2 material. The clause outlines that this could be done by:

a) engaging with other industry participants, eSafety and other interested stakeholders on different age assurance options through government-led technology trials and consultation processes

b) collaborating on the development of improved national and international approaches to age assurance and related issues such as the use of children's data, privacy preservation, data security, accessibility and respect for the best interests of children

c) collaborating with domestic and international regulators, NGOs, industry associations and other peak bodies and stakeholders in activities of the kind referred to in paragraphs (a) and (b)

d) actively engaging in the Age-Restricted Material Codes review process.

eSafety considers this list is not exhaustive and encourages industry participants to improve reliability, robustness and effectiveness of age assurance through other means if available to a service.

Testing, monitoring, and improving age assurance measures over time

Certain service providers are also obligated to take steps to **test, monitor, and (in some cases) improve their age assurance measures over time**, including:

- Social media services (compliance measures 7.1 and 10.1)
- Relevant electronic services (compliance measures 7.1, 7.2 and 16.1)
- Designated internet services (compliance measures 7.1, 9.5, 10.1 and 10.19)
- Search engine services (compliance measure 7.23(b)).

Testing age assurance measures

As noted above, service providers should ensure any age assurance method employed has undergone sufficient testing and evaluation before use and **while in use**. Service providers are encouraged to **undertake and document ongoing internal testing procedures as well as seek external audits or independently validated testing**, and provide that information in code compliance reports as required.

Monitoring age assurance measures

eSafety considers that service providers should conduct regular monitoring, including by maintaining awareness of:

- **changes in circumvention methods** and associated risks – for example, where generative AI may be used for fraudulent documents or to attempt to bypass facial age estimation
- **changes in end-user behaviour and demographics**, including where children and young people **migrate to different services** where they experience different risks and harms. In this instance any insights should be provided to eSafety
- **community expectations of privacy**
- **scams, privacy complaints and data breaches** that may emerge in response to increased uptake of age assurance
- **new developments in age assurance**, including exploring new and emerging methods that are more privacy preserving or require less end-user data without sacrificing accuracy.

Service providers are required to share the findings and outcomes of their reviews and evaluations with eSafety on request as required in code compliance reports. This includes documenting any changes made in response to lessons learned during early implementation.

Improving age assurance over time

Service providers should regularly review their implementation of safety measures and update them where appropriate, especially where new approaches better support the online safety objectives of the Age-Restricted Material Codes, including by:

- utilising findings gathered from testing and monitoring age assurance methodologies to innovate and update their age assurance methods
- investing in appropriate systems, tools and processes to support continuous improvement of age assurance measures
- considering interoperable options, digital wallet integrations and zero-knowledge proof methods, that decrease end-user burden and enable control over personal information

- incorporating additional sources of age information as they become available, such those shared from a device or app store
- participating in initiatives such as support for research, pilot projects, and collaboration with non-government and government organisations or cross industry collaboration which aim to improve age assurance methods.

Reasonable steps to prevent circumvention

Clause 5.1(c) of the Head Terms to the Age-Restricted Material Codes notes that service providers should take reasonable steps to ensure that their age assurance measures cannot be circumvented.[125]

Service providers should maintain awareness of circumvention methods that may be used by end-users. This includes methods of **identity- or age-based circumvention** or **location-based circumvention** such as:

- using VPNs
- relying on other age verified end-users
- creating false identity documents
- using AI or deepfakes to spoof age estimation systems
- clearing cache or browser history to reset age checks
- answering knowledge-based questions with guessed or known information
- using hand-me-down devices that retain age assured end-user credentials.

Examples of reasonable steps service providers may take to ensure that their age assurance measures cannot be circumvented

- Ensuring age assurance measures incorporate liveness checks
- Using device telemetry, behavioural signals, and other persistent identifiers to detect irregular activity
- Integrating VPN detection services and IP intelligence APIs to flag and restrict high-risk IP ranges
- Using geolocation consistency checks to identify mismatches between IP address and declared location

[125] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl 5.1(c)(i).

- Implementing systems to detect and investigate suspicious IP switching
- Considering additional signals that can indicate likely location

Review of age assurance decisions

As detailed above, clause 5.1(c)(i) of the Head Terms to the Age-Restricted Material Codes requires service providers to consider fairness in applying appropriate age assurance measures. eSafety considers that a relevant consideration here is that service providers should offer accessible and timely complaints or review mechanisms for end-users in relation to:

- access control measures decisions
- any adverse outcomes resulting from any age assurance processes.

Review of access control measure decisions

Service providers should clearly communicate how and when end-users can make a complaint or seek review of a decision they believe was made in error.

These mechanisms should be **accessible and inclusive**, allowing end-users to navigate the process **with clarity and relative ease**. These mechanisms should be clear and readily identifiable to end-users at the point a decision is made about their access to controlled material or services.

When requesting additional information as part of this process – such as identification – **service providers should not require end-users to provide government-issued identification material without also providing a reasonable alternative means for end-users** to assure the service provider that they are not under the age of 18. As detailed above, eSafety considers that an age assurance mechanism where the use of government identification documents is the sole method to verify a user's age will likely not satisfy the principle in clause 5.1(c)(i) of the Head Terms to the Age-Restricted Material Codes that service providers should consider fairness in applying appropriate age assurance.

End-users should be notified when their complaint or application for review is received, along with an expected timeframe for a response. When receiving a response to a review or complaint, the end-user should also be informed of how their complaint, dispute or request for review was assessed – whether by a human moderator or an AI system, and the reason for the outcome.

In alignment with best practice approaches to artificial intelligence and automated decision-making, there should be human in the loop or human oversight to mitigate the risks of incorrect decisions. **Fully automated reviews should be avoided**.

Review of complaints relating to adverse outcomes resulting from any age assurance processes

As required under the Age-Restricted Material Codes, certain service providers also have obligations to provide end-users with a mechanism to provide feedback to the service in relation to the service provider's compliance with the Age-Restricted Material Codes, including their age assurance obligations. In certain instances, service providers are also required to refer unresolved complaints to eSafety. Further information on these obligations can be found in **Part 3 of the Regulatory Guidance**.

Service providers should ensure they are monitoring and recording relevant metrics and indicators of end-users' experiences in making complaints disputes or requesting review. For example, if a high number of successful complaints or reviews are being made about a particular tool or technology, this can be a useful indicator of the effectiveness and performance of that tool or technology. Service providers are also encouraged to make it clear to end-users that they can make privacy complaints to the OAIC.

By way of further example, if a significant number of complaints or reviews are made by end-users about one tool, technology or process – but not others – this can be an indicator that end-users are not able to find or use the mechanism to make complaints, disputes or request review for those other tools, technologies or processes.

Service providers are well placed to determine what metrics should be monitored and tracked in relation to complaints, disputes and reviews, and should ensure they are able to report to eSafety when required to do so.

Other requirements

Appropriate policies, procedures, systems and technologies

Clause 5.1(e) of the Head Terms to the Age-Restricted Material Codes notes that service providers should supplement age assurance obligations with appropriate policies, procedures, systems and technologies. This will help ensure that age assurance systems are transparent and clear to end users.

This could be done by providing information about service providers' use of age assurance and other measures. Service providers should use age-appropriate language and be

accessible to people of different literacy levels and abilities, including children. It should include:

- plain-language explanations of when and why age assurance is required
- guidance on what age assurance options are being used or are available to users
- what personal information will be collected, used, how and where it will be stored and protected, possible outcomes, and what the service provider will do with the result – including what is retained or destroyed and what other privacy protections are in place and relevant transparency obligations under the Privacy Act
- information about how users can seek review of a decision.

Service providers should be able to report on the uptake of their support resources, make them easy to access for all end-users and promote their availability to the public.

Transparency measures are also key to building trust and addressing the community's concerns about age assurance. eSafety understands that there is low public awareness among both Australian adults and children of the range of age assurance technologies available, including those currently in use, and how they work in practice.[126] Participants in the consumer research commissioned by DITRDCSA also reported low trust in services and held concerns about the privacy and security of their information, which can result in a lower willingness to engage with certain age assurance measures.[127]

Terms of use, standards of conduct, policies and procedures

Terms of use, standards of conduct, policies and procedures are key mechanisms for service providers to communicate what is and is not allowed on their services and the minimum age for end-users of their service. **Service providers should ensure that terms of use, standards of conduct, policies and procedures** for the service, and information available to end-users include:

- whether class 1C and class 2 material is permitted on the service
- the age assurance processes in use on the services
- the data collection, use and retention policy
- relevant review processes, including in relation the age assurance process reviews, or content review processes.

[126] Department of Infrastructure, Transport, Regional Development, Communications, Sports and the Arts, *Age assurance consumer research findings* (18 June 2025) p 22.
[127] Department of Infrastructure, Transport, Regional Development, Communications, Sports and the Arts, *Age assurance consumer research findings* (18 June 2025) p 22.

Steps to prevent children from accessing or being exposed to class 1C and class 2 material

If the age assurance measures implemented by a service provider indicate that an end-user is a child, the service provider is required to take steps to prevent them from accessing or being exposed to class 1C and class 2 material on their service. These steps differ dependent on the applicable Age-Restricted Material Code.

Table 2 of this Appendix provides a summary of these obligations contained within the Age-Restricted Material Codes.

Table 2: Summary of steps required to be taken by service providers after age assurance measures indicate that an end-user is a child

Relevant Age-Restricted Material Code	Steps required to be taken after age assurance measures indicate that an end-user is a child to prevent children from accessing or being exposed to class 1C and class 2 material
Internet Search Engine Services	Implement safety measures for account holders that are likely to be Australian children (compliance measure 7.2)
App Distribution Services	Prevent Australian children from downloading or purchasing adult apps (compliance measure 7.1)
Designated Internet Services	Implement access control measures on relevant services (compliance measures 7.1, 9.5, 10.1 and 10.19)
Relevant Electronic Services	Implement access control measures on relevant services (compliance measures 7.1 and 16.1)
Social Media Services (Core Features)	Implement access control measures on relevant services (compliance measures 7.1 and 10.1)

Access control measures

Some service providers are required to implement access control measures to put into effect the findings of any relevant appropriate age assurance measures. eSafety notes that these access control measures may be different from those required under other schemes like the RAS Declaration.

Service providers can implement access control measures either:

- before providing access to the service, or
- before providing access to the relevant class 1C or class 2 material.

What is appropriate will depend on the service and the risk of access or exposure to that material on a service. In particular, eSafety considers that:

- it would be appropriate for providers of services with the sole or predominant purpose of enabling access to high-risk material to implement access control measures before providing access to the service
- where a service elects to serve high-risk material to end-users via a personalised algorithmic feed, even if not all material served may be high-risk, this indicates that implementing access control measures before providing access that feature would be appropriate
- where the design of a service means that child end-users can be reasonably prevented from accessing or being exposed to class 1C and class 2 material while retaining access to the remainder of the service, access control measures may be reasonably limited to those parts of the service.

Reporting on compliance with age assurance obligations

As outlined in **Part 3 of the Regulatory Guidance**, certain service providers will be required to submit compliance reports outlining what steps they have taken in order to comply with the requirements under the Age-Restricted Material Codes.

Providers of services which are obligated to implement age assurance measures should ensure that they are prepared to provide information about:

- standard operating procedures or other systems, processes or policies in place underlining how age assurance is undertaken
- steps taken to test, monitor and/or improve age assurance measures over time
- the number of complaints and appeals related to age-eligibility decisions for each age assurance method used, and the time taken to process these complaints/appeals
- steps taken to address circumvention.

Importantly, eSafety will **not** request from service providers any information that identifies end-users, including any government-issued identification documents.

Table A: Age assurance requirements under the Age-Restricted Material Codes

Code	Services required to implement appropriate age assurance	Requirements to monitor age assurance over time?**	Outcome if age assurance returns an under 18 result
Internet Search Engine Services Code Commences 27 June 2026*	Compliance measure 7.2: All services, for logged in account holders	✓ Compliance measure 7.23(b): test, monitor and improve	Compliance measure 7.2: Implement safety measures for account holders likely to be Australian children.
Social Media Services (Core Features) Code Commences 9 March 2026	Compliance measure 7.1: Services that allow online pornography, self-harm material, or high-impact violence material	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 10.1: Tier 1 AI companion chatbot features provided as part of a social media service	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 10.2: Tier 2 AI companion chatbot features provided as part of a social media service if safety by design defaults not implemented	✓ Test and monitor	Implement access control measures on relevant services
Relevant Electronic Services Code Commences 9 March 2026	Compliance measure 7.1: Services with the sole or predominant purpose of permitting end-users to share online pornography and self-harm material	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 7.2: Gaming services that are, or would likely be, classified as R18+ under the Classification Act (including simulated gambling material)	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 16.1: Tier 1 AI companion chatbot features provided as part of a relevant electronic service	✓ Test and monitor	Implement access control measures on relevant services

	Compliance measure 16.2: Tier 2 AI companion chatbot features provided as part of a relevant electronic service if safety by design defaults not implemented	✓ Test and monitor	Implement access control measures on relevant services
Designated Internet Services Code Commences 9 March 2026	Compliance measure 7.1: Tier 1 services	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 9.5: Classified DIS when providing access to X18+ and/or simulated gambling material	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 10.1: Tier 1 high impact generative AI DIS before the generation of high-risk material	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 10.2: Tier 2 high impact generative AI DIS before the generation of high-risk material if safety by design defaults not implemented	✓ Test and monitor	Implement access control measures on relevant services
	Compliance measure 10.19: Tier 1 high impact generative AI DIS before providing access to high-risk material	✓ Test and monitor	Implement access control measures on relevant services
App Distribution Services Code Commences 9 September 2026*	Compliance measure 7.1: services with a high or medium risk that adult apps (those rated 18+) will be downloaded or purchased from the service	✗	Prevent Australian children from downloading or purchasing adult apps

*** Age assurance requirements only. Other measures come into effect 6 months prior.**

**** All service providers have obligations for ongoing work on age assurance under clause 1.3 of the Head Terms to the Age-Restricted Material Codes.**

Appendix G: Suicide, self-harm and eating disorder material under the Age-Restricted Material Codes

Acknowledgements

eSafety acknowledges the meaningful and expert contributions of representatives from the following Australian agencies and organisations:

- National Suicide Prevention Office
- National Taskforce for Social Media and Eating Disorders
- Orygen
- Butterfly Foundation
- ReachOut
- headspace
- Kids Helpline
- Beyond Blue

Sources of support

This guidance contains information and descriptions about eating disorders, self-harm, and suicide, and may be upsetting or distressing.

Where possible this material has been discussed sensitively and only in the level of detail required. This regulatory guidance is targeted towards service providers.

If you experience, or if anyone you know is experiencing distress or requiring additional support, please ensure that you reach out to support services within your respective location.

Introduction

Under the Age-Restricted Material Codes, service providers have several obligations in relation to **self-harm material**. This Appendix provides guidance to help service providers comply with their obligations in relation to this material.

In developing this regulatory guidance, eSafety undertook a targeted consultation in December 2025 with services and individuals with expertise in the areas of suicide, self-injury, and eating disorders (see the Acknowledgements section).

eSafety's Self-Harm Material Consultation Summary (**consultation summary**) outlines several further matters raised during the consultation which service providers could take into account when considering how best to approach self-harm material for the purposes of the Age-Restricted Material Codes. In some cases, some of the matters in the consultation summary go beyond the requirements of the Online Safety Codes and Standards, given the current legislative framework of the Online Safety Act and National Classification Scheme. Nevertheless, eSafety encourages service providers to consider these matters as they provide useful insights from subject-matter experts and those with relevant lived experience.

What is self-harm material?

Under the Head Terms to the Age-Restricted Material Codes, **self-harm material** is defined as:

> a subcategory of class 2 material defined for the purposes of [the Codes] as being comprised of **material** that is class 2 material because it **encourages, promotes or provides instruction** for:
>
> a) **suicide**;
>
> b) an act of deliberate **self-injury**; and/or
>
> c) an **eating** disorder or **behaviour associated with an eating disorder**.

Under the Head Terms, where relevant, service providers must develop a process for categorising unclassified material in a way that is informed by the National Classification Scheme.[128]

To assist service providers, Tables 1a and 1b of this Appendix provides guidance on key terms in relation to self-harm material under the Age-Restricted Material Codes.

[128] Consolidated Industry Codes of Practice for the Online Industry (Class 1C and Class 2 Material) Head Terms, cl. 3(f).

This guidance is intended to:

- assist with the interpretation of key terms used in the codes
- provide context where these terms are not defined in the National Classification Scheme
- promote international regulatory coherence, aligning where possible with other international jurisdictions who regulate similar content.

This reflects the findings of our consultation with subject matter experts who highlighted the need for global consistency in these definitions to ensure that service providers adequately action this material at scale on what are mostly global online services.

eSafety is also a member of the Global Online Safety Regulators Network and maintains a strong commitment to consistency and collaboration in online safety regulatory frameworks to ensure that online services can benefit from economies of scale when targeting this material.[129]

Table 1a: Guidance on key terms included in the definition of self-harm material under the Age-Restricted Material Codes

Term	Guidance
Suicide	An action that a person takes to deliberately end their own life.
Self-harm/self-injury[130]	An act in which a person intentionally harms themselves typically as a way of dealing with emotional distress, painful internal experiences, or overwhelming situations. An intentional act is an act that someone undertakes willingly or knowingly regardless of motive or intended outcome.
Eating disorder	Eating disorders are serious, complex and potentially life-threatening mental illnesses, which impact upon a person's physical health, mental health and holistic wellbeing. They are characterised by disturbances in behaviours, thoughts and feelings towards body weight and shape, and/or food and eating.
Behaviours associated with eating disorders	This includes behaviours that are generally associated with eating disorders across the various presentations. This non-exhaustive list includes the following behaviours: caloric restriction or fasting, binge eating, purging or other compensatory behaviours, avoidance or restriction of food, and excessive exercise.

[129] eSafety Commissioner, The Global Online Safety Regulators Network.
[130] While the wording in the Age-Restricted Material Codes refers to 'self-injury', we recognise that the recommendations from consulted subject matter experts – based on extensive engagement to develop the 'National Suicide Prevention Strategy 2025-2035', it was identified that 'self-harm' material would be the preferred terminology. Given the established naming conventions in the Age-Restricted Material Codes, to avoid confusion we use the term 'self-harm/self-injury' to ensure this can be appropriately distinguished from the broader definition whilst also recognising this as the preferred terminology.

Table 1b: Guidance on key terms in relation to types of material referenced in the definition of self-harm material under the Age-Restricted Material Codes

Term	Guidance
Material that promotes self-harm	Material which publicises, supports or recommends suicide, self-harm, or eating disorders (and associated behaviours). This also may include material that promotes the concealment or 'masking' or any form of self-harm. It is not required for this promotion to be intentional or explicit. This includes content which glamourises, glorifies, romanticises or normalises self-harm.
Material that encourages self-harm	Material which could incite or persuade others to contemplate or engage behaviours associated with suicide, self-injury, eating disorders, or behaviours associated with an eating disorder and/or make others more likely to attempt or consider this as a course of action. Encouragement does not have to be intentional or explicit. This can include material which glamourises, glorifies, romanticises, or normalises self-harm.
Material that provides instructions for self-harm	Describes or depicts a method or any actions that may be instructive in nature for suicide, self-injury/self-harm, eating disorders, or behaviours associated with an eating disorder, in sufficient detail that it can be emulated or replicated. Material does not need to deliberately or explicitly provide instructions for self-injury, suicide or eating disorders to be harmful. Instructions may be minimal and still able to be emulated. This could include describing or showing visually the materials or actions that might be undertaken that constitutes self-harm, such as setting out a 'plan' or steps, showing a diagram, providing 'tips', 'coaching' or 'guidance'.

Material will only be **self-harm material** for the purposes of the Age-Restricted Material Codes where it is class 2 material.[131] The impact and context of material will influence whether a piece of content is classified as class 2 material, and accordingly whether the material would be considered as self-harm material under the Age-Restricted Material Codes.[132]

Relevant to the terms 'Material that promotes self-harm', 'Material that encourages self-harm' and 'Material that provides Instructions for self-harm' outlined in Table 1b of this Appendix, it is unlikely that material that solely provides information about treatment, support, or help-seeking behaviours related to suicide, self-injury or eating disorders will meet the definition of self-harm material. **eSafety's position is that Australian end-users should not be inhibited from accessing credible and non-harmful resources that:**

- **provide information about self-harm**

[131] Please see the section titled 'What material is covered by the Codes and Standards?' for more detail on when material may be considered class 2.
[132] Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, Australian Classification, How a rating is decided.

- **provide information about treatment or support available**
- **promote help-seeking behaviour.**

Higher harms associated with AI-generated self-harm material

Material generated by artificial intelligence (AI) chatbots is interactive in nature. Under the National Classification Scheme, services that are interactive in nature are more likely to be higher impact, and therefore higher risk.[133] Providers of generative AI services which are interactive, such as AI companion chatbots, should be aware that self-harm material generated by their services is more likely to be high impact (see also discussion in Appendix E).

Obligations related to self-harm material

Table 2 of this Appendix provides a summary of obligations that services have in relation to self-harm material under the Age-Restricted Material Codes.

Table 2: Summary of obligations for service providers in relation to self-harm material under the Age-Restricted Material Codes

Code	Implement safety tools	Detect and remove self-harm material	Provide safety information	Engage with trust and safety providers
Internet Search Engine Services Code	✓	×	✓	✓
Social Media Services (Core Features) Code	✓ (Services which allow self-harm material)	✓ (Tier 1 and 2 services which do not allow self-harm material)	✓	✓ (Tier 1 and 2 Services)
Relevant Electronic Services Code	×	×	✓	✓
Designated Internet Services Code	✓ (Services which allow self-harm material)	✓ (Tier 1 and 2 services which do not allow self-harm material)	✓ (Tier 1 Services)	✓ (Tier 1 Services)
App Distribution Services Code	✓	×	✓	✓

[133] *Guidelines for the Classification of Films 2012*, part 2.

This following section provides information about how service providers can meet their obligations to:

- implement safety measures for self-harm material
- detect and remove self-harm material
- provide safety information
- engage with trust and safety providers.

Implementation of safety measures for self-harm material

Service providers are required to comply with obligations that relate to the implementation of safety measures to prevent end-users from accessing or being exposed to self-harm material. Table 3 of this Appendix provides a summary of these obligations.

Table 3: Summary of services required to implement default safety measures for self-harm material

Code	Obligations	Risk tier/category of service required to comply
Internet Search Engine Services Code	**Compliance measure 7.6: Search advertising** Providers must take appropriate steps to ensure that advertising for self-harm material is not served to account holders that the provider knows with reasonable certainty is an Australian child.	All search engine services
	Compliance measure 7.10: Measures to reduce unintentional exposure to self-harm material A provider of an internet search engine service must apply measures to protect and prevent end-users from being unintentionally exposed to self-harm material. At minimum this includes protections that promote trustworthy content, and measure to prevent autocomplete predictions that result in search queries seeking self-harm material.	All search engine services
	Compliance measures 7.11 & 7.12: Crisis prevention material A provider of an internet search engine must imply means to detect and provide crisis prevention information in response to search queries regarding suicide, deliberate self-injury and eating disorders.	All search engine services

Code	Obligations	Risk tier/category of service required to comply
App Distribution Services Code	**Compliance measure 7.5: Implementation of safety tools,** which may include a feature that provides safety information to Australian end-users who search for adult apps (including crisis prevention material).	All app distribution services
Designated Internet Services Code	**Compliance measures 7.2, 10.2 & 10.20: Continuous improvement for systems regarding self-harm material** A provider of a service that does not allow self-harm material must take steps to continuously improve systems which can detect and action self-harm material before it is encountered by end users.	Tier 1 and tier 2 DIS and high impact generative AI DIS that do not allow self-harm material.
	Compliance measures 7.5 & 10.21: Safety tools Implementation of safety tools for all end-users to limit access or exposure to self-harm material.	Tier 1 and tier 2 DIS that allow self-harm material and high impact generative AI DIS that allows access and exposure to self-harm material
	Compliance measure 10.13: Location on or via service that provides online safety information, including information about how Australian end-users can contact counselling and support services.	Tier 1 high-impact generative AI DIS
Social Media Services (Core Features) Code	**Compliance measure 7.2**: **Implementation of safety tools,** for all end users to limit access or exposure to self-harm material.	Social media services that allow self-harm material.

Safety tools to limit access or exposure to self-harm material

A number of services have obligations to implement safety tools in relation to limiting access or exposure to self-harm material and activating them by default for child end-users. Some compliance measures in the Age-Restricted Material Codes note that this may include implementing several different kinds of solutions.[134]

[134] See for example, Social Media Services (Core Features) Code (Age-Restricted Material), cl 7.2.

Some of these solutions are set out in the following list, with accompanying examples of supportive practices as raised by experts during the consultation (see the consultation summary for more details):

- Filtering and blocking self-harm material.
    - Adult end-users should be able to opt-in and out to this functionality. Where possible, they should be able to select what material they want to be filtered, blocked or blurred.
- As noted in the consultation summary, experts consider this should ideally include providing 'shield functions', for more information refer consultation summary.[135]
- Removing self-harm material from recommended content (including targeted advertisements).
    - For example, service providers could give the users the ability to opt-out of targeted advertisements (particularly for appearance-related advertising) and that for children as a default this is should not be enabled.
    - This could also include allowing end-users to reset their recommender algorithms easily and on demand. Services should also provide clear and accessible information about the effect and duration of the algorithmic reset to users.
- Blurring self-harm material or placing interstitial notices on self-harm material so that users can click through to view if they wish.
- Halting autoplay of self-harm material.
    - This could also extend to other features that are intended to encourage user engagement. For example, infinite scroll functionality might appropriately be disabled if detected that the recommender system is returning a high volume of results relating to self-harm material.
    - Services should provide end-users with clear and accessible information about the risks associated with using these features, particularly in relation to self-harm material.

Service providers are required take steps to ensure and evaluate the effectiveness of these safety tools where they are implemented. eSafety may seek information about the effectiveness of safety tools when assessing compliance with relevant measures in the codes.

[135] For further information, see page 33 of the consultation summary.

To increase the effectiveness of safety tools, service providers should do the following:

- Ensure that safety tools are **easy to use, accessible, and developmentally appropriate.**
- Keep end-users regularly **informed** about any safety tools. For example:
    - Informing end-users of all safety tools available to them and how they can be activated, including when these safety tools are updated.
    - When changing safety settings, end-users should be provided with relevant warnings as it pertains to health and safety.
- **Meaningfully engage with mental health professionals and experts** throughout the product development lifecycle so that safety tools related to self-harm material are evidence-based and effective.

Detecting and removing self-harm material

Under the Age-Restricted Material Codes, certain service providers will be required to take steps to detect and remove self-harm material. Table 4 of this Appendix provides a summary of these obligations.

Table 4: Summary of services required to detect and action self-harm material

Code	Obligations
Social Media Services (Core Features) Code	✓ Compliance measure 8.2 (Tier 1 and tier 2 social media services which do not allow self-harm material)
Relevant Electronic Services Code	✓ Compliance measure 10.2 (Dating services, if against terms and services)
Designated Internet Services Code	✓ Compliance measure 7.2 (Tiers 1 and Tier 2) ✓ Compliance measure 10.20 (high impact generative AI DIS)

Concerns have been raised that automatic detection and actioning of harmful material, including self-harm material, may inadvertently lead to the 'over-capture' of beneficial information and support services.[136]

eSafety's position is that Australian end-users should not be inhibited from accessing credible and non-harmful resources that:

- **provide information about self-harm**
- **provide information about treatment or support available**
- **promote help-seeking behaviour.**

[136] Joint Select Committee on Social Media and Australian Society, Social media: the good, the bad, and the ugly, November 2024.

eSafety notes that there are several steps that service providers may be required to take under the Age-Restricted Material Codes which can lower the risk of over-capture of credible and non-harmful material. For example:

- Service providers have obligations to allow and action complaints from end-users, as outlined in Table 7 of this regulatory guidance.
    - When doing so, service providers should use information from reports to monitor emerging trends in how self-harm material may present itself on these services. This reflects evidence that self-harm material can quickly evolve and develop over time.
- Several classes of service providers are required to take steps to continuously improve the detection of self-harm material on their services over time, as discussed in the next section.

Continuously improving the detection of self-harm material

Several classes of service providers are required to take steps to continuously improve the detection of self-harm material on their services over time. Table 5 of this Appendix provides a summary of these obligations.

Table 5: Summary of services required to continuously improve their detection of self-harm material over time

Code	Requirements to continuously improve detection of self-harm material
Social Media Services (Core Features) Code	✓ Compliance measure 8.2 (Tier 1 and tier 2 social media services which do not allow self-harm material)
Designated Internet Services Code	✓ Compliance measure 7.2 (Tier 1; Tier 2) ✓ Compliance measure 10.20 (High impact generative AI DIS)
Internet Search Engine Services Code	✓ Compliance measures 7.23(h) and 7.23(j)

Appendix D of this document outlines the steps that service providers should take in order to comply with their continuous improvement obligations under the Age-Restricted Material Codes.

The consultation summary identifies the views of experts that service providers should also consider ongoing and meaningful engagement with crisis services, lived-experience groups and academic resources. This is to ensure service providers are aware of and able to action emerging trends related to self-harm material.

Provision of information in relation to self-harm material

Various services have obligations to provide information and guidance to end-users to mitigate risks associated with self-harm material. These obligations are summarised in Table 6 of this Appendix.

However, as a matter of community safety, eSafety recommends **all services** provide their end-users with appropriate crisis and mental health information and services where practical to do so – particularly if service providers have identified that an end-user may be at risk of harm.

Table 6: Summary of services required to provide support information in relation to self-harm material

Code	Requirements to provide support information in relation to self-harm material
Search Engine Services Code	✓ Compliance measures 7.11 and 7.12: Provide crisis prevention information in response to general queries regarding suicide, an act of deliberate self-injury, eating disorders or behaviours associated with an eating disorder.
App Distribution Services Code	✓ Compliance measure 7.5: Implement safety tools, which may include a feature that provides safety information to Australian end-users who search for adult apps.
Relevant Electronic Services Code	✓ Compliance measures 8.11, 9.11, 10.12, 11.11 & 15.10: Provide information about steps end-users can take to manage or mitigate risks relating to self-harm material. ✓ Compliance measures 8.12, 9.12, 10.13, 11.12, 12.1 & 15.11: Provide help seeking guidance – including counselling and support.
Social Media Services (Core Features) Code	✓ Compliance measure 9.13: Provide help seeking guidance – including counselling and support.
Designated Internet Services Code	✓ Compliance measures 7.16 & 10.13: Provide help seeking guidance – including counselling and support.

Crisis prevention material

Under the Search Engine Services Code (Age-Restricted Material), service providers are required to provide crisis prevention information, including:[137]

- a helpline associated with a reputable organisation that is able to provide support relevant to children

[137] Internet Search Engine Services Code (Age-Restricted Material), cl 7.11 and 7.12.

- link(s) to information and support that is freely available and relevant to children through a reputable organisation.

The consultation summary contains further discussion about what experts said service providers should consider when providing support information to end-users.

A list of **Australian providers** can be found at Appendix A of the consultation summary. This list reflects the point-in-time feedback from the consultation related to the production of this Appendix and is not intended to be exhaustive.

Support information in relation to AI-generated content

Obligations in relation to the provision of support information outlined in Table 6 of this Appendix also extends to AI-generated self-harm material, including material generated by generative AI tools and AI companion chatbots.

When providing support information, providers should consider the specific risks associated with AI-generated self-harm material and ensure that end-users are made aware of these risks. This includes the risk of inaccuracies in AI-generated material (particularly when end-users are seeking medical advice in relation to self-harm related behaviours).

Provision of support information should be done clearly, regularly, and in a manner that is developmentally appropriate, and may include:[138]

- regularly reminding users that AI features or services are non-human
- ensuring transparency on how AI systems are designed, in a way that is understandable to young users and empowers them to make informed choices how they use AI systems
- stating clearly any limitations (including the potential inaccuracy) in AI-generated material by:
    - warning users that AI-generated information is not an appropriate substitute for professional health advice or intervention,
    - reminding end-users to contact a human professional, and
    - referring appropriately to verified sources of health information.

[138] American Psychological Association, Artificial Intelligence and adolescent well-being: An APA Health Advisory, 2025.

Engagement with safety and community organisations

Certain services are required to engage with safety and community organisations to inform measures taken to prevent children from accessing or being exposed to class 2 material, including self-harm material, under the Age-Restricted Material Codes. Table 7 in this Appendix has a summary of these obligations.

Table 7: Summary of services required to engage with safety and community organisations

Code	Services required to engage with safety and community organisations
Designated Internet Services Code	✓ Compliance measures 7.22, 10.10 (Tier 1 DIS; high-impact generative AI DIS)
Social Media Services (Core Features) Code	✓ Compliance measure 9.11
Social Media Services (Messaging Features) Code	✓ Compliance measure 10.15
Relevant Electronic Services Code	✓ Compliance measure 8.15, 9.15, 8.16, 11.15, 15.15 (Closed communication RES; other communication RES; dating services; gaming service with communications functionality; tier 1 RES)
App Distribution Services Code	✓ Compliance measure 7.11
Internet Search Engine Services Code	✓ Compliance measure 7.22

Working with reputable safety and community organisations will help to ensure that services are providing contemporary, evidence-informed and contextually relevant information and resources to children. The consultation summary provides insights from service providers and community organisations about how service providers may effectively engage with them.[139]

[139] See for example, page 30 of the consultation summary.

eSafety.gov.au