COMMUNICATIONS ALLIANCE LTD

Communications Alliance Submission

to the Office of the eSafety Commissioner

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

and

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

and ancillary materials:

Discussion Paper: Draft Online Safety Standards

Fact Sheet: Draft Online Safety (Designated Internet Services Class 1A and Class 1B Material) Industry Standard 2024

Fact Sheet: Draft Online Safety (Relevant Electronic Services Class 1A and Class 1B Material) Industry Standard 2024

25 January 2024

CONTENTS

Communications Alliance

Communications Alliance is the primary communications industry body in Australia. Its membership is drawn from a wide cross-section of the communications industry, including carriers, carriage and internet service providers, content providers, platform providers, equipment vendors, IT companies, consultants and business groups.

Its vision is to be the most influential association in Australian communications, co-operatively initiating programs that promote sustainable industry development, innovation and growth, while generating positive outcomes for customers and society.

The prime mission of Communications Alliance is to create a co-operative stakeholder environment that allows the industry to take the lead on initiatives which grow the Australian communications industry, enhance the connectivity of all Australians and foster the highest standards of business behaviour.

For more details about Communications Alliance, see http://www.commsalliance.com.au.

Introduction

Communications Alliance welcomes the opportunity to make a submission to the Office of the eSafety Commissioner in response to the exposure drafts of the *Draft Online Safety (Relevant Electronic Services – Class 1A and 1B Material) Industry Standard 2024* (RES standard) and the *Draft Online Safety (Designated Internet Services – Class 1A and 1B Material) Industry Standard 2024* (DIS standard) (jointly, the standards), the associated Discussion Paper and Fact Sheets.

Communications Alliance and its members take their roles in relation to the safety of users online very seriously. All members invest substantial amounts of resources and time into systems, processes and/or technologies that aim at the reduction of harms arising from online scams and materials.

Consequently, we also wholeheartedly support the aim of minimising harm emanating from the dissemination of and access to Class 1A and B material online, and we look forward to further constructive engagement with the Office of the eSafety Commissioner (eSafety) and other relevant stakeholders on the development of practical industry standards for these important and complex industry sections.

Working with this submission / context:

Several Communications Alliance members have substantially contributed to the submission prepared by the Digital Industry Group Inc. (DIGI). Accordingly, we do not seek to replicate all issues expertly raised or the detail provided in DIGI's submission but, instead, support DIGI's submission. Therefore, this submission ought to be read in close conjunction with and in addition to the issues and arguments advanced in DIGI's submission.

Where appropriate, we provide additional feedback and/or indicate where we see a need for alternative commentary.

In addition to this submission, we are providing a marked-up version of the exposure drafts of the standards that contain additional comments that may not be reflected in this submission. Usually these relate to language and drafting suggestions. Please be aware that the marked-up versions of the exposure drafts contain a very limited set of comments on specific points only and do not contain all changes proposed to relevant provisions, or to the Standards more generally, by this submission and DIGI's submission. They should therefore be read in that light.

We will not be explicitly addressing the questions raised in eSafety's Discussion Paper as many of the issues raised in those have previously been elaborated on in discussions between industry and eSafety. However, the feedback provided in our submission will invariably also include comments in relation to the questions posed in the Discussion Paper.

We note that feedback provided in relation to Class 1A material will often also be valid for the proposed drafting of requirements in relation to Class 1B material, even if this may not be expressly mentioned.

Similarly, while we attempt to identify whether feedback pertains to the proposed RES or DIS standard, or both, it will be useful for eSafety to cross-check either standard for similar issues.

Members may also provide their own submissions.

1. Complexity of draft standards

1.1. Having drafted codes for RES and DIS ourselves, we appreciate the difficulties associated with the task of drafting the corresponding standards. Many of the complexities result from underlying definitions of the *Online Safety Act 2021* (OSA) and the breadth of the industry sections to be covered.

However, we fear that the proposed standards are too complex for most users. Industry experts, who were involved in the drafting of the codes, struggle to fully understand the proposed standards, in particular the DIS standard. Smaller organisations or those not involved in the codes process and other stakeholders, e.g. consumer representatives and end-users affected by the standards, will encounter even greater challenges. We also see a range of areas where the approach taken is inconsistent with the OSA (see for example, comments on definitions below).

We recommend revisiting the structure and definitional aspects of the draft standards to enhance simplification and consistency with the OSA.

2. Approach to minimum compliance measures

2.1. Pages 10 and 11 of the Discussion Paper note four key considerations that eSafety has taken into consideration in the drafting of the standards. These are:

- *"the importance of striking a balance between flexibility and enforceability, so service providers, eSafety and third parties have clarity about required outcomes*
- *the principle of risk-based, outcomes-based and technology neutral regulation so providers can implement measures that reflect the characteristics of their service and are responsive to rapidly shifting technologies*
- *ensuring obligations are meaningful as well as technically feasible, practical and – where appropriate – able to be deployed at scale*
- *the importance of human rights, including the right to freedom of expression, the right not to be subjected to arbitrary or unlawful interference with privacy, the right to protection from exploitation, violence and abuse, and the rights and best interests of children."*

2.2. These considerations are critical to the effective operation of the standards, but none are expressly included in either the proposed RES or DIS standard. We strongly recommend these four points be reflected in the standards themselves (or at least guidance that we believe ought to be provided by eSafety) as underlying principles for the operation of the standards.

2.3. Unfortunately, in our view, several of the proposed measures do not align easily with the practicality consideration (i.e. whether providers can implement the measures and whether the measures reflect the characteristics of their service) or with aspects of the human rights considerations, including privacy. (Naturally, we would expect the standards not to require unlawful actions.) We urge eSafety to reconsider the standards on the basis of the feedback received during the consultation process and in light of the four considerations it has highlighted as informing the standards.

We note in particular that privacy considerations are expressly acknowledged in similar overseas legislation such as in Recital 47 of the *EU Digital Services Act* and in section 22 of the *UK Online Safety Act 2023.* Importantly, the current *draft EU Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse* does not permit the scanning of end-user-managed hosting services of EU data subjects without a *"targeted, specified and limited"* legal order based on *"reasonable grounds of suspicion"*.

2.4. As part of eSafety's reconsideration of these issues, one way to assist with reflecting some of these points would be to reintroduce relevant elements from section 6 of the

Head Terms to the registered codes, and the balancing test in section 5.1(b) of the Head Terms to the codes.

3. Application of draft standards

3.1. The draft standards propose a 'predominant functionality' test in order to determine whether a standard applies or, instead, whether another industry standard or code apply to a respective RES or DIS.

3.2. This test does not align with the test applied in the Head Terms of the registered industry codes which instead apply a 'predominant purpose' test.

3.3. The OSA also relies on purpose as a key criterion of the definition of social media services, the industry section likely to require most assessment in relation to overlap of services with the RES standard.

3.4. It is not clear why eSafety has chosen to deviate from the concept used in the codes and, at this stage, we do not see any rationale or merit in doing so. A test relying on predominant functionality is, in our view, infeasible.

3.5. The inconsistency between the registered industry codes and the standards as drafted will lead to substantial difficulties to determine which industry standard or code ought to be applicable to the service under consideration, noting that both industry standards and codes seek to apply only one standard/code to the same RES/DIS.

3.6. Consequently, to aid understanding and compliance, we strongly recommend that the standards apply the same test as already used in the registered industry codes.

4. Definitions

Definitions relating to material

4.1. The definitions for various types of materials are unnecessarily complex in language and we recommend revision.

4.2. It is imperative that all definitions that relate to various categories of Class 1A and 1B material as well as the definition of Class 1A and 1B material themselves reference the Classification Scheme and associated legislative texts instead of seeking to replicate some, but not all, elements of the Classification Scheme.

4.3. Section 106 of the OSA equally ties back to the *Classification (Publications, Films and Computer Games) Act 1995*, as also do the already registered industry codes.

4.4. Importantly, the draft standards oversimplify the approach to the respective materials and deviate from the approach of the Classification Scheme in that the proposed definitions do not reference the requirements of the Classification Scheme to consider certain material, including pro-terror and drug-related material, in its context or the context in which it appears online. We note that the *Guidelines for the classification of Computer Games 2012, Guidelines for the Classification of Films 2012* and *Guidelines for the Classification of Publications 2005* clearly refer to the "*crucial importance of context*". The definition of 'justification' in the standards may be intended to align to the concept of context but does not adequately replicate the approach taken in the Classification Scheme including its Code and Guidelines.

4.5. We request that the definitions for all materials reference, but do not repeat, the Classification Scheme in full to ensure that the standards are consistent with the Scheme, the OSA and the registered industry codes. Otherwise, significant compliance challenges will arise for industry.

4.6. In addition, we fear that future changes to the Classification Scheme – which is currently under review – will not be reflected in the standards but would apply to the OSA and registered codes.

4.7. We note that the ACMA Guidance on Code development states: "*However, while codes are required to be both based upon, and consistent with, the Telecommunications Act and instruments made under that Act, they should not repeat or paraphrase it. One of the main roles of codes is to provide industry-initiated solutions to issues that are not covered by legislation.*" [emphasis added] We do not see any reason why the same rationale would not apply to regulator-drafted industry standards.

4.8. It is worth noting that the definition of Class 1B material and other definitions as they relate to drug-material in combination with the requirements to assess material in context (or in relation to a 'justification') have the potential of creating compliance issues in view of the global deregulation of some drugs, including the decriminalisation of marijuana and other drugs in Australia.

4.9. In addition, we are concerned that the proposed drafting does not appear to allow for use of alternative language, that may be more end-user-friendly, to describe the respective materials. This again deviates, in our view unnecessarily, from the approach of the registered codes and ought to be remedied.

End-user and enterprise RES/DIS

4.10. From our experience with the registered codes so far, we recommend including a clarification (in the standards themselves) as to what a natural person is, i.e. not a business. This is one of the most commonly asked questions that the industry associations have received from smaller and medium-sized organisations. It will also assist stakeholders understand the scope of the standards.

4.11. Conversely, the definition of enterprise RES/DIS ought to omit the brackets "*(and not an individual)*" to avoid confusion with sole traders who, while being an 'organisation', often are also acting as individuals. We believe the expression in brackets is unnecessary for the definition but prone to creating confusion.

Definition of 'provide'

4.12. The reliance of the definition of 'provide' on the making available of a service introduces fundamental problems. As noted in a number of items of this submission, a 'provider' of a RES/DIS often may not have control over the technology that is being on-sold or used in the 'making available' of the respective service. The reference to the 'provider' of the services as the entity that makes the service available, therefore, leads to requirements that these entities cannot comply with.

RES standard-specific definitions

Impractical functionality concept

4.13. Further to our commentary in relation to the use of 'functionality' in the applicability test, the reliance of functionality in the definition of sub-categories for RES does not reflect operational reality nor does it address, so we believe, the intended outcome. For example, the ability to create lists is not a primary functionality of RES but rather an ancillary function. The ability to communicate directly with the chosen party is the primary 'function' or, as we would say, the primary purpose of the service.

Closed communication RES

4.14. The draft standards define closed communication RES inclusive of SMS/MMS services provided by carriage service providers (CSPs). This contradicts the Fact Sheet that accompanies the draft standard which states that email services provided by CSPs are supposed to be included in the definition but not SMS/MMS services. This ought to be rectified.

4.15. However, we note overall our objection to including services provided by CSPs into the definition of closed communication RES and, consequently, pre-assessed RES.

4.16. Many of the obligations subsequently imposed on pre-assessed RES are, so we believe, either technically, operationally and/or legally infeasible or disproportionate. Consequently, alternative drafting arrangements ought to be pursued to exclude these services from those obligations. This includes services provided by CSPs but also RES provided by other providers, e.g. end-to-end encrypted services and services where providers do not enter into a contractual/customer relationship with all participants of the supply chain (also refer to our comments at item 11 below).

It is not helpful to seek to impose obligations on CSPs/other RES with view to requesting that these organisations subsequently explain the infeasibility of the application of the requirements for their respective organisations. This creates an unnecessary regulatory burden and also confusion for other stakeholders, including consumers.

4.17. It is also important to understand that the vast majority of CSPs (or telephony RES and closed communication RES as currently defined in the draft standard) are not carriers themselves but resellers. These organisations usually have very limited control over most technical aspects associated with the service that they are reselling.

4.18. We note that the TIO advised that there are currently 1,700 CSPs registered with the TIO, and the ACMA has advised that there are currently 1,000 CSPs with connected services in the Integrated Public Number Database (IPND).

4.19. Further, the requirements ought to be amended to apply only where the risk of harm emanating from the material being accessible on a RES is proportionate (having regard to whether the requirement sought to be imposed is appropriately tailored to achieve the end(s) sought) to the regulatory and financial burden associated with compliance with the requirement (and where compliance is technically, operationally and/or legally feasible, without an explanation/justification requirement) and the impact(s) of the requirement on end-users (including end-user privacy, security and freedom of expression).

Telephony RES

4.20. Following on from the above, the definition of telephony RES ought to be amended to reference all RES provided by CSPs (and include a reference to the definition of CSP in section 87 of the *Telecommunications Act 1997*). The definition of closed communication RES ought to be amended to exclude CSP RES.

4.21. Importantly, but in addition to the feasibility constraints (purely technical or otherwise) noted at item 5 below, CSPs do not necessarily share eSafety's understanding that the blanket monitoring required to fulfil the envisaged detection, deterrence, removal etc. obligations can be authorised through the standard as envisaged by eSafety. Also, to the extent the required measure necessitated the interception of communications, this would be prohibited under Chapter 2 of the *Telecommunications (Interception and Access) Act 1979*. If eSafety has received specific legal advice on this matter, it would be helpful to advise of the source and share the advice with Communications Alliance.

Open communication RES

4.22. The definition of open communication RES *"includes a relevant electronic service that enables an end-user to invite, through use of an internet link, another end-user to communicate with the first end-user."* While we believe we understand the intention of the drafting, the proposed language would include an email or other message into which a link is included inviting a user to communicate. We recommend amending the definition to reflect the actual intention.

DIS standard-specific definitions

Definitional complexity

4.23. We expect that the online section of DIS is the section that covers most online services. At the same time, the definitions proposed for DIS in the draft DIS standard are so complex that we do not believe they will be able to be understood by a large majority of such services which, usually, do not have any legal assistance or training. We recommend revisiting the definitional constructs, including the number of definitions for different DIS and cross-references.

Definitions for AI/ML-related DIS

4.24. Overall, the definitions for Artificial intelligence (AI) and machine learning (ML)-related DIS take insufficient account of the fact that in many instances the provider (as defined in the standard) of the respective DIS has no or limited control over the technical features of the service it makes available.

Enterprise services

4.25. Similar issues arise in relation to enterprises. The DIS standard (and to some extent the RES standard) suggests that enterprise customers are 'providers' of services to their employees. However, such enterprise customers do not have the technical capability or control over the technical features of the services in the same manner as the actual (original) provider of the service.

We provided more specific drafting feedback in the marked-up exposure draft of the standard. Also refer to our comments at item 13.6 further below.

This issue requires further careful thought, and we welcome a discussion with eSafety.

5. Technical feasibility

Criteria of feasibility

5.1. In our view, the proposed drafting is too narrow and does not appropriately reflect the considerations that ought to apply to feasibility – whether technical or otherwise.

5.2. We also note a feasibility test ought to be consistently applied to all elements of the standards that require systems, processes and/or technology to be applied, content to be reviewed, removed etc.

5.3. Similarly, it appears that technical (or other) feasibility does not factor into an assessment of the 'appropriateness' of an action. This ought to be remedied, e.g. by including 'technical feasibility' (including the revised considerations indicated below) into the 'appropriate' test. We also refer to our feedback to include all considerations of section 5.1(b) in the Head Terms of the registered codes into definition of appropriate. Please also refer to DIGI's submission on this issue.

5.4. The drafting currently focuses on costs as the key criterion as to whether a certain action ought to be taken. However, a number of other considerations are equally important and ought to be added to the list of items that ought to be considered when assessing feasibility. These include, for example:

- the commercial availability of technical solutions, and the efficacy and viability of such solutions;
- whether an action would require the provider to implement or build a systemic weakness, or a systemic vulnerability (including a new decryption capability or actions that would render systemic methods of authentication or encryption less effective) into a form of electronic protection, or would prevent the provider from rectifying a systemic weakness, or a systemic vulnerability, in a form of electronic protection. (Note: this reflects the language of section 317ZG of the *Telecommunications Act 1997*.);
- whether the provider of the service is capable to access, view and/or remove individual items of material communicated via the service and has sufficient visibility of content and end-user activity to determine, in context, the nature of the material (as being Class 1A/B) and whether a breach of terms of service etc. has occurred (We suggest consideration be given to including the concepts of 'capable of reviewing and assessing materials' and 'capable of removing materials' into the standards, in line with the drafting in the proposed codes.);
- the level of control a provider has which will impact its ability to take action;
- whether the relevant action can be taken in line with global standards and protocols applicable to the relevant service type;
- the proportionality of the action in relation to the likelihood of risk of harm emanating from the service; and
- human rights (e.g. freedom of expression) and privacy interests of the end-users using the service.

5.5. The *UK Online Safety Act 2023* and the *EU Digital Services Act* provide useful guidance as to other criteria that ought to be included for consideration.

5.6. We highlight that it is important to include a consideration into the standards themselves that puts beyond doubt that any building or weakening of security measures will not be required of providers, similar to the approach that was taken by the Australian Government in the *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018*.

5.7. Communications Alliance CSP members note that many of the requirements in relation to SMS, MMS and CSP email services, as currently drafted in the RES standard, are not feasible, either because of physical technical limitations and/or because the implementation of measures would be vastly disproportionate the likely harm caused and exceedingly costly to implement.

5.8. CSP email systems run on networks and systems that were not designed to provide these services. They are ancillary to the services of internet access and the provision of a mobile/fixed network. Many have been built to global standards, past or still applicable. Consequently, these networks and systems are far less adjustable (i.e. there are no simple 'bolt-ons' or network upgrades that could be used). CSPs have informed us that the requested measures would most likely require a 're-build' of systems associated with multi-year change programs and leading to unmanageable costs.

5.9. In line with our comments at items 4.14 to 4.19, we believe that the core requirements of the standards applicable to RES ought to reflect these realities from the outset.

5.10. Non-CSP members hold similar concerns for other RES that may be end-to-end encrypted or where the contractual relationship does not extend through the supply chain.

5.11. Given our comments above on additional criteria for the assessment of feasibility (i.e. commercial availability of solutions, the need to maintain the security of networks and systems, a provider's capability to assess, review and remove material, control, consideration of global standards, proportionality, and human rights) we recommend omitting the word 'technical' and simply using the term 'feasibility'.

Education

5.12. In 2022, DIGI and Communications Alliance commissioned research to provide an evidence-base of the expectations of the general Australian public in relation to online safety. This research showed, for example, substantial concerns by the public with the scanning of personal devices, private storage and one-to-one messages.[1]

5.13. If eSafety indeed proceeds with the proposed requirements that will involve such scanning of communications and similar measures (which eSafety should reconsider for the reasons outlined in these submissions), we urge eSafety to give consideration to undertaking a broad, high-visibility educational campaign with the aim of educating the Australian public about the new requirements and the reasons for those.

6. Risk assessments and methodology

Consideration of existing mitigations

6.1. The assessment of risk in section 9 of both standards assumes that service providers do not inherently take measures that reduce risk. Instead, the section appears to assume that an assessment of risk occurs 'in a vacuum', without regard to existing mitigations. A risk assessment should not just focus on the potential for abuse based on the inherent functionality of the service – instead, the extent to which a provider has already implemented appropriate risk mitigations should be relevant to a service's risk profile.

Missing purpose criterion for DIS risk assessments

6.2. In the DIS standard, the central role of 'purpose' proposed in the industry DIS code has not found its way into the language of the proposed standard. However, industry believes that the criterion of 'purpose' in key to any risk assessment due to the broad range of DIS. As a result, the proposed DIS standard is likely to create significant uncertainty as to the applicable risk tiers. This ought to be rectified.

Risk assessment criteria

6.3. Section 8(1) of the draft RES standards, as currently drafted, is impractical and/or disproportionate: arguably any service could be used, at some point in time, for the distribution etc. of the material in consideration. As drafted, a single instance of this occurring would trigger the assessment requirement. Even worse, as drafted in future tense, i.e. "*will*", the best available interpretation is that a risk assessment is always

[1] "Headline findings of the research include:
- 89% did not believe that the Code should cover all the digital services being considered, and they are particularly sensitive to the scanning of personal devices, private storage and one-to-one messages.
- 80% believed that there should be a suspicion of possessing or sending restricted material before scanning is performed, with 78% agreeing that a warrant should be in place first.
- 79% did not find all the potential consequences of detection of restricted materials (deletion, suspension or reporting) acceptable, with 59% preferring that detected materials is flagged with warnings instead.
- 71% did not think that all categories of restricted content should be scanned for either, and where they are searched 65% stated that the technology should be 100% accurate before being used.
- 45% disagreed with the basic premise, preferring that restricted content not be scanned for.
- Many are confused on what is illegal and restricted, and differ in expectations of what should be covered."

required. This would mean, for example, that SMS/MMS require a risk assessment. In our view, these framing issues mean that the risk assessment 'gating criteria' are not meaningful, and this ought to be remedied. One approach could be to link the requirement for risk assessment to a risk of material harm arising from the service being used in such a way.

Requirement of risk assessment for CSPs

6.4. Following on from the above and our comments at item 4.20 and 4.21, section 8(6) of the RES standard lists categories of RES to which the requirements to undertake a risk assessment do not apply. CSP services (as redefined; or currently telephony RES and assuming SMS/MMS are included in closed communication RES) ought to be included in this list as the risk profile for those services is the same, and they are not Tier 1 services.

Matters to be taken into account:

6.5. The list of matters to be taken into account in the RES standard (section 9(5)) also ought to include, as proposed for the DIS standard, *"the manner in which material is created or contributed to in connection with the service"*. This would assist with appropriately considering context-specific communications.

6.6. Section 9(5)(g) of the RES standard and 9(5)(j) of the DIS standard limit safety by design guidance or tools to a regulator or international body. However, safety by design approaches ought also to be permissible if derived from individual organisations and technology services. We recommend amending these sections accordingly.

Reporting on risk assessments:

6.7. Section 33 of the RES standard applies to all RES. However, not all RES are required to conduct a risk assessment. The section ought to be amended to reflect this.

A similar issue arises for section 40 of the DIS standard in relation to DIS that are not required to perform a risk assessment.

7. Terms of use

7.1. Section 14 of the RES standard and other subsequent sections refer to terms of use for a particular service. However, some services, such email services provided by CSPs and SMS and MMS which are part of a mobile service offering, do not have separate terms of use and/or terms of use for the associated mobile service would not commonly be read by users looking for the information in question. Consequently, the provisions of these sections ought to be drafted to refer so that 'terms of use for the service' is specifically defined to include acceptable use policies and/or standard terms and conditions.

7.2. We also note the various criminal offences established by section 474 of the *Criminal Code Act 1995* (in relation to using a carriage service to distribute CSAM, suicide related material, abhorrent violent material and the general offence to menace, harass or cause offence), i.e. the distribution of certain material may already be prohibited under law.

7.3. Our comments here are in addition to those raised in the DIGI submission.

8. Awareness of material and subsequent actions

8.1. Section 14(3) requires providers, upon awareness, to enforce "*contractual rights in respect of the breach in an appropriate way that is proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach.*" The qualification ought to be augmented to also include consideration of the extent to which a risk of harm may arise from the interruption of the service and from incorrectly identifying a breach, including (but not limited to) false positives.

8.2. Not all providers and/or services will have equal capabilities and it is important to weigh the likelihood of 'getting it wrong' and the harm from the loss of a service with the potential harm emanating from the material under consideration.

8.3. It is unclear as to how awareness of a breach will be established for either Class 1A or Class 1B material. Does eSafety envisage that a RES provider acts to apply the required actions once an external party has determined a breach of terms or similar, or is it envisaged that, e.g. upon a complaint, a RES provider must have the ability to determine whether a breach has occurred?

8.4. If the latter is being intended, how would a provider that is unable to review material on their services, such as SMS/MMS services or encrypted services, determine a breach? In this case, the provider would need to rely on the party making a complaint/assertion and accept at face value that the material presented to the provider has been communicated using the service.

8.5. This appears unworkable for some services, especially given the severe consequences that may result, such as the suspension of a mobile phone service as SMS/MMS services cannot be suspended/terminated independent of the mobile services provided to the end-user.

8.6. We note that mobile numbers are frequently 'spoofed', e.g. the actual mobile number that is associated with a specific service is over-stamped (or 'spoofed') with a number belonging to another person and subsequently used to make calls or send texts.

8.7. Note that this is only one possibility of bringing an innocent person potentially in breach of terms of use (with subsequent enforcement measures). We are aware of other ways of deliberately inflicting enforcement action onto an end-user (without account hacking or similar), without that end-user actually having breached terms of use.

8.8. Similar issues arise in other sections of the standards that rely on awareness of a breach.

8.9. It is also not clear what would be expected of providers with regards to the requirements to review reports by end-users where such providers have no ability to view or access the content in question. It may also not be possible to review content (and hence provide much assistance to a reporting person) where content has been encrypted.

8.10. Generally, it should be understood that for many services, the removal of the material in question will not be possible because the communications are not stored after transmission. Where it is possible, removal of communications from a communications service will only result in the removal of specific instances of the material from the service, i.e. the material will remain available to the end-user when stored locally on a device. The expected effectiveness of removal of material from a service on its availability ought to be considered in the proportionality of measures that are requested to be applied by providers.

9. Requirements in relation to (known) pro-terror material

9.1. As per numerous previous discussions, we do not share the view that a database of known pro-terror material exists that RES providers could use to detect and remove known pro-terror material.

9.2. We also re-iterate the need for contextual analysis required to assess whether communications of such material could be justified. Such analysis is not possible at scale and/or without substantial human review (which most communications providers are not qualified to undertake). Reliance on artificial intelligence to undertake this task is, at this stage, inappropriate and likely to result in high numbers of false positives due to the immaturity of this technology.

9.3. We equally do not know of any authority that could verify such material against the requirements of the OSA/Classification Scheme (or the RES standard for that matter), i.e. as against Australian legal requirements.

9.4. We refer to DIGI's submission for an in-depth discussion of this (and related issues).

10. Development programs

10.1. The draft RES Standard (section 23) and the draft DIS standard (section 24) require certain types of RES/DIS to have a development program subject to a threshold of 1 million or more average monthly active end-users of the service, in Australia, over the immediate previous financial year.

The 1M threshold equates to about 4% of Australia's population. We suggest eSafety align with the EU and UK in determining such a threshold and use (as in those jurisdictions) a 10% population threshold, i.e. about 2.5 million average monthly active end-users in Australia.

Please also note that in the case of DIS, not all providers of the captured services may be able to capture the location of their end-users, thereby making the application of this threshold difficult or impractical.

10.2. Our comments here are in addition to those raised in the DIGI submission and apply to the extent the development program obligations are retained.

11. In-service concept (RES Standard s27(3) and DIS Standard s29(2)(b)) and complaints

11.1. The RES standard uses an in-service concept for reporting, complaints and the provision of information/tool mechanisms etc. in several requirements.

This concept does not apply or is unworkable for some services, such as SMS/MMS or email provided by a CSP. For other types of services, it is unclear how this concept is meant to operate and may not be operationally useful, e.g. apps do not have a separate website; or what constitutes 'in-service' for an email service which may often be accessed through a third-party app or service (e.g. the email app on a smart phone)?

Consequently, we request alternative language be used, e.g. information be provided, or reporting/complaints be able to be received in a manner that is clear and accessible to end-users of the service in Australia.

11.2. Related to the in-service concept is the notion that the communication being used to distribute Class 1A/B material is occurring within the realm of one RES provider (i.e. the same service provider has a relationship with the sender and recipient of the communication) and/or that the RES provider receiving the communication is able to

determine its source. However, this is not the case for SMS and MMS, e.g. the receiving CSP will not necessarily be able to determine which network the SMS/MMS originates from and, consequently, is unable to progress a complaint in a meaningful way. Again, similar problems arise for all email services, e.g. where individualised domains (e.g. me@randomname.com) have been used. Also, as previously noted, all services that lack a contractual/customer relationship with end-users of other providers for a relevant communication face similar issues.

11.3. It also worth noting that providers of these services are already subject to the ACMA-made *Telecommunications (Consumer Complaints Handling) Industry Standard 2018* and subject to many prescriptive requirements in relation to any complaint associated with their service. Against this background, if it is deemed that complaint handling requirements for CSPs are indeed required in the RES standard, such requirements for CSPs ought to be assessed by eSafety for consistency with the Complaints Handling Standard.

12. Reporting obligations

12.1. Much of the information that providers are required to report is likely to be commercially highly sensitive in nature. The standards are lacking provisions that would protect the confidentiality of the information provided to eSafety.

We request that the confidentiality protections provided in the Head Terms (at section 7.3(b)) be included in the standards.

13. AI / Machine learning in DIS standard

Inclusion of AI /ML into the DIS standard

13.1. We acknowledge eSafety's desire to address evolving risks associated with generative AI through the draft DIS standard, and support eSafety's stated intention of ensuring any obligations are proportionate and targeted.

13.2. It is important to understand that the potential risks arising from the use of AI, particularly generative AI, cannot be considered in isolation. The entry of many different business models for making machine learning models and AI technology available to a variety of users into the market has led to increased regulation over this technology – a trend that is to continue at pace for at least another year. Globally, governments are wrestling with privacy, security, safety, trade, intellectual property, and other concerns through their legislative process.

13.3. In the Australian context, the Department of Industry, Science and Resources (DISR) is leading a whole-of-government process to develop regulatory mechanisms to ensure AI is used safely and responsibly. In addition, there are other related Australian Government initiatives, such as the review of Australia's privacy, cybersecurity and copyright regimes. In our response to DISR's Discussion Paper *Safe and Responsible AI in Australia*, we articulated that to date, AI has been capably regulated through existing technology-neutral legislation and regulation. We also stated that any gap analysis should take into account whether existing legislation or regulation was capable of appropriately assigning responsibility across the AI supply chain, and whether unlawful and illegal use of AI applications was appropriately prevented or mitigated by existing legislation or regulation.

13.4. It is critical that eSafety's approach to addressing online safety risks posed by AI is consistent with the Australian Government's approach (as recently articulated), as well as international standards. We are concerned that the proposed requirements in the DIS standard would prematurely introduce definitions that do not align with developing global standards and are difficult to understand and apply to real-world business

models. We encourage eSafety to engage with DISR as part of this process, and caution against the introduction of highly specific language in industry standards that is separate from this overarching regulatory process, thereby risking the creation of potentially contradictory or competing priorities.

13.5. We are also of the view that the DIS standard does not need to specifically regulate the distribution of AI as the primary ways models are distributed are covered already by either provisions in relation to enterprise DIS providers (when AI technologies are offered on an enterprise basis), or the tiered DIS provider approach. To the extent an AI service is offered on a business-to-consumer basis, the consumer AI service would undergo a risk assessment to determine the relevant tier like any other site or service. Allowing companies the flexibility to undertake risk assessments (using prescribed guardrails, and with eSafety empowered to gather information) acknowledges the nascent and developing state of AI technologies, such as generative AI, and reduces the risk of developing a patchwork of laws that do not align with the purpose of the Australian Government in regulating AI.

Practical difficulties with the obligations on AI / ML providers as proposed

13.6. If eSafety insists on pre-defined categories and on limiting a service provider's ability to undertake independent risk assessments for their services, we make the following specific observations:

- Generally, the DIS standard does not account for the fact that the organisations who are 'providers' (also note our comments regarding the definition of this term at item 4.12) of AI/ML services often do not have control over the technology as they are customers of other technology providers. As a consequence, the draft standard contains numerous obligations that cannot realistically be complied with by these organisations.

  Consider, for example, an enterprise DIS that makes an AI application available to its enterprise customers (e.g. Microsoft CoPilot). The enterprise DIS will not be able to control the application or the content generated as it cannot modify it. The ability to control the AI sits with the real provider of the AI as opposed to the organisation that made it available. The generation of content will be undertaken by the end-users of the customer organisation. Likewise, the obligations associated with the category of 'machine learning model platform service' (MLMPS) are largely impossible to comply with..

  Page 7 of the Fact Sheet to the DIS standard (and p. 23 of the Discussion Paper), eSafety notes "*machine learning model platform services are not capable of reaching into the models themselves […] platforms can and do moderate what models they distribute*" [emphasis added]. An MLMPS can control which models they select for the service, but once a machine learning model is downloaded, the MLMPS offering it has no engagement with the downloader, the content or the end-users of the model (if externally deployed).

  An MLMPS has no control over the end-user or application of the model once it is downloaded, and cannot be expected to comply with obligations that are not practically possible, such as responding to end-user reports, engaging with or presenting information about eSafety to end-users, or scanning any content such as prompts or outputs. Of the obligations listed in the DIS standard, an MLMPS likely can only comply with sections 1, 14 and 38(2) to (6).

- Consequently, question 15 from the Discussion Paper in relation to the drafted requirements being "*best-placed to prevent the use of generative AI features to create and disseminate class 1A and class 1B material*" has to be negated.

- We suggest the DIS standard should focus on the key differentiating factors, i.e. who controls or deploys the content to the end-user and has the relationship with the end-user.
- For this reason, the DIS standard should:
  - in section 13 (potential overlaps) clarify that determining which code or standard applies will entail an assessment of whether the respective provider deploys or controls the content or service as opposed to the functionality of a service; or
  - either merge the definition of MLMPS with enterprise DIS or ensure the obligations for MLMPS providers mirror the obligations for enterprise DIS (note that neither an MLMPS nor an enterprise DIS can comply with section 23(2)).

14. Implementation timeframe

14.1. Section 2 of the draft standards envisages an implementation timeframe of 6 months. While this timeframe was also included (industry had sought a longer timeframe) in the registered industry codes, we note that implementation of measures for RES and DIS are likely to be more complex than for many other services. The complexity inherent in the two industry sections is, in part, reflected in the fact that eSafety and industry were unable to agree on codes that fulfilled eSafety's requirements in respect to appropriate community safeguards.

14.2. Depending on the compliance measures in the final industry standards, a compliance timeframe of 12 months or longer may be required to implement the systems, processes and/or technologies envisaged for the detection and removal of material.

14.3. At the very least, the arrangements as provided in section 7.1(b) of the Head Terms[2] ought to be replicated in the draft standards to ensure that RES and DIS providers have an opportunity, in good faith, to work towards compliance with complex requirements.

15. Conclusion

Communications Alliance looks forward to continued engagement with eSafety and all relevant stakeholders in developing these important standards.

We continue to lend our support to meaningful measures that will assist with the minimisation of harms arising from the distribution and access of Class 1A and B materials online.

For any questions relating to this submission please contact Christiane Gillespie-Jones on 02 9959 9118 or at c.gillespiejones@commsalliance.com.au.

[2] Section 7.1(b) of the Head Terms to the registered industry codes:

"If after the effective date in section **Error! Reference source not found.** eSafety notifies an industry participant that it is non-compliant with a measure required under this Code and the participant has reasonable grounds for not being fully compliant, the participant will not be in breach provided that it can demonstrate to eSafety's reasonable satisfaction that it is working towards achieving compliance on or before the first anniversary of the date of registration.

Note: Examples of reasonable grounds for not being fully compliant by the date specified in section **Error! Reference source not found.** may include circumstances where significant engineering or system changes are required in order to implement a measure."

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

I, Julie Inman Grant, eSafety Commissioner, determine the following industry standard.

Dated

DRAFT ONLY—NOT FOR SIGNATURE

Julie Inman Grant
eSafety Commissioner

Contents [table of contents removed]

Part 1—Preliminary

1 Name

This is the *Online Safety (Designated Internet Services—Class 1A and Class 1B Material) Industry Standard 2024*.

2 Commencement

This industry standard commences on the day that is 6 months after the later of:

(a) the day after the day on which it is registered under the Act; and

(b) the day after the day on which it is registered under the *Legislation Act 2003*.

3 Authority

This industry standard is determined under section 145 of the *Online Safety Act 2021*.

4 Object of this industry standard

The object of this industry standard is to improve online safety for Australians in respect of class 1A material and class 1B material, including by ensuring that providers of designated internet services establish and implement systems, processes and technologies to manage effectively risks that Australians will solicit, generate, distribute, get access to or be exposed to class 1A material or class 1B material through the services.

Commented [A1]: Throughout both standards, this reference and similar references should always include "and/or". It is impossible and also not necessary for all providers captured to provide all of the above. This comment applies wherever these terms appear.

5 Application of this industry standard

(1) This industry standard applies to a designated internet service, wherever it is provided from, but only so far as it is provided to end-users in Australia.

Commented [A2]: The extraterritorial application of the OSA is already dealt with in the OSA, including through the definition of Australians. It is unnecessary and confusing to insert additional qualifiers in the standards. If absolutely necessary, include in guidance.

(2) If:

(a) this industry standard applies to a designated internet service; and

(b) another industry standard, or an industry code, applies to the service; and

(c) the service's predominant functionality is more closely aligned with the other industry standard or the industry code;

this industry standard does not apply to the service.

Commented [A3]: Refer to our and DIGI submissions. At the very minimum this should reference "functionality **or** purpose"

Part 2—Interpretation

Note: A number of expressions used in this industry standard are defined in the Act, including the following:

(a) child;
(b) class 1 material;
(c) class 2 material;
(d) Classification Board;
(e) Commissioner;
(f) computer game;
(g) consent;
(h) designated internet service;
(i) material;
(j) parent:
(k) posted;
(l) publication;
(m) removed;
(n) service.

6 General definitions

Definitions

(1) In this industry standard:

acceptable use policy, for a designated internet service, means the provisions of the terms of use for the service that regulate the use of the service by end-users.

account holder, for a designated internet service, means the person who is the counterparty to the agreement for the service for the provision of the service.

Example: A designated internet service may be provided to a family, where a parent or carer has the agreement with the provider of the service. The parent or carer is the account holder. Family members (including the parent or carer who is the account holder) who use the service are end-users.

Act means the *Online Safety Act 2021*.

appropriate action: see section 12.

Australian child means a child who is in Australia.

child sexual abuse material or ***CSAM*** means material that:

(a) describes, depicts, promotes or provides instruction in child sexual abuse; ~~or~~

~~(b) is known child sexual abuse material.~~

child sexual exploitation material or ***CSEM*** means material that:

(a) is or includes material that promotes, or provides instruction in, paedophile activity; or

(b) is or includes:

(i) child sexual abuse material; or

(ii) exploitative or offensive descriptions or depictions involving a person who is, appears to be or is described as a child; or

Commented [A4]: Refer to our and DIGI submissions with respect to all definitions that relate to material/Classification Scheme.

Commented [A5]: Known CSAM is defined separately.

(c) describes or depicts, in a way that is likely to cause offence to a reasonable adult, a person who is, appears to be or is described as a child (whether or not the person is engaged in sexual activity);

and, in the case of a publication, also includes material that is or includes gratuitous, exploitative or offensive descriptions or depictions of:

(d) sexualised nudity; or

(e) sexual activity involving a person who is, appears to be or is described as a child.

class 1A material means class 1 material so far as it comprises:

(a) child sexual exploitation material; or

(b) pro-terror material; or

(c) extreme crime and violence material.

Note: For the definition of ***class 1 material*** see section 106 of the Act.

class 1B material means class 1 material so far as it comprises:

(a) crime and violence material (but not extreme crime and violence material); or

(b) drug-related material.

Commented [A6]: Refer to our submission.

Note: For the definition of ***class 1 material*** see section 106 of the Act.

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

Note: RC is a classification.

classified DIS means a designated internet service that has the sole or predominant purpose of providing general entertainment, news, or educational content, being:

(a) films or computer games that:

(i) have been classified R18+ or lower; or

(ii) are exempt from classification in accordance with the *Classification (Publications, Films and Computer Games) Act 1995*; or

(b) films or computer games that have not been classified but, if classified, would likely to be classified R18+ or lower; or

(c) books, newspapers and magazines, whether in digital or audio form, podcasts and or digital music that, if required to be classified, would likely to be classified Category 1 or lower;

and includes a service that is taken to be a classified DIS under subsection 13(2).

compliance report means a report required by section 38.

crime and violence material, in relation to a computer game, means material that is accessible in a computer game and that, without justification:

Commented [A7]: "material that is a computer game" does not make sense.

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in, or promotion of, matters of crime or violence; or

(b) is or includes depictions of bestiality or similar practices; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes depictions of violence that:
  (i) have a very high degree of impact; and
  (ii) are excessively frequent, prolonged, detailed or repetitive; or

(e) is or includes depictions of cruelty or realistic violence that:
  (i) have a very high degree of impact; and
  (ii) are very detailed; or

(f) is or includes depictions of actual sexual violence; or

(g) is or includes depictions of implied sexual violence related to incentives or rewards.

crime and violence material, in relation to a publication, means material that is, or is included in, the publication and that, without justification:

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in matters of crime or violence; or

(b) is or includes realistic depictions of bestiality; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes gratuitous, exploitative or offensive descriptions or depictions of violence that:
  (i) have a very high degree of impact; and
  (ii) are excessively frequent, emphasised or detailed; or

(e) is or includes gratuitous, exploitative or offensive descriptions or depictions of cruelty or real violence that:
  (i) have a very high degree of impact; and
  (ii) are very detailed; or

(f) is or includes gratuitous, exploitative or offensive descriptions or depictions of sexual violence.

crime and violence material, ~~in relation to a material~~ that is not a computer game or a publication, means material that, without justification:

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in matters of crime or violence; or

(b) is or includes depictions of bestiality or similar practices; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes gratuitous, exploitative or offensive depictions of violence that:
  (i) have a very high degree of impact; or

(ii) are excessively frequent, prolonged or detailed; or

(e) is or includes gratuitous, exploitative or offensive depictions of cruelty or real violence that:

(i) have a very high degree of impact; and

(ii) are very detailed; or

(f) is or includes gratuitous, exploitative or offensive depictions of sexual violence.

designated internet service has the meaning given by section 14(1) of the Act.

development program means a program required by section 24.

DIS means a designated internet service.

drug means a chemical, compound, or other substance or thing, that is included in Schedule 4 of the *Customs (Prohibited Imports) Regulations 1956.*

drug-related material, in relation to a computer game, means material that, without justification:

(a) depicts the unlawful use of drugs in connection with incentives or rewards; or

(b) depicts interactive, detailed and realistic use of drugs, being unlawful use; or

(c) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or

(b) is or includes detailed instruction in the unlawful use of drugs.

drug-related material, in relation to a publication, means material that, without justification:

(a) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or

(b) is or includes detailed instruction in the unlawful use of drugs.

drug-related material, ~~in relation to material~~ that is not a computer game or a publication, means material that, without justification:

(a) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or

(b) is or includes detailed instruction in the unlawful use of drugs; or

(c) is or includes material promoting the unlawful use of drugs.

end-user, of a designated internet service, means a natural person who uses the service.

Commented [A8]: Refer to our submission.

Example: A designated internet service may be provided to a family, where a parent or carer has the agreement with the provider of the service. The parent or carer is the account holder. Family members (including the parent or carer who is the account holder) who use the service are end-users.

end-user managed hosting service means:

(a) a designated internet service that is primarily designed or adapted to enable end-users to store or manage material; and

(b) includes a service that is taken to be an end-user managed hosting service under subsection 13(2).

Note 1: Examples of end-user managed hosting services include online file storage services, photo storage services, and other online media hosting services, including such services that include functionality to allow end-users to post or share content.

Note 2: For purposes of this industry standard, an enterprise DIS that meets this definition will be taken to be both an enterprise DIS and an end-user managed hosting service – see subsection 13(2)(c).

Note 3: An end-user managed hosting service differs from third-party hosting services (as defined in the Hosting Services Online Safety Code (Class 1A and Class 1B Material)) which have the sole or predominant purpose of supporting the delivery of another service online and which do not directly interact with end-users.

enforcement authority means:

(a) a police force or other law enforcement authority; or

(b) an organisation (including a non-government organisation) the functions of which include receiving reports of child sexual exploitation material or pro-terror material and facilitating making those reports to law enforcement authorities.

enterprise customer means the account holder under the agreement for the provision of an enterprise DIS.

Note: The enterprise customer will often make the service available to a class of end-users, such as its staff.

enterprise DIS means a designated internet service:

(a) the account holder for which is an organisation ~~(and not an individual)~~; and

(b) the primary ~~functionality~~ purpose of which is to enable the account holder, in accordance with the terms of use for the service, to use the service for the organisation's activities, including integrating the service into the organisation's own services that are or may be made available by the organisation to the organisation's end-users; and

(c) that is of a kind that is usually acquired by account holders for the purpose mentioned in paragraph (b);

and includes a service that is taken to be an enterprise DIS under subsection 13(2).

Note 1: An enterprise DIS excludes Third-Party Hosting Services as (as defined in the Hosting Services Online Safety Code (Class 1A and Class 1B Material) and which are dealt with by that Code).

Note 2: An enterprise DIS would, for example, include:

(a) websites designed for the ordering of commercial supplies by enterprise customers; and

(b) services which provide pre-trained artificial intelligence or machine learning models for integration into a service deployed or to be deployed by an enterprise customer.

Commented [A9]: Refer to comment below re s 13(2)(c) why this does not make sense for just EUMH and should be extended for all overlapping enterprise DIS use cases (i.e. when any DIS is offered on an enterprise basis).

Commented [A10]: Refer to our submission.

Commented [A11]: It is unclear why functionality is relevant here. The purpose of offering the service is the key question. The functionality will be the same whether it is being offered to a natural person, or an organisation.

exploitative, in relation to a description or depiction of an event, means that the description or depiction:

(a) appears intended to debase or abuse, for the enjoyment of readers or viewers, the person or entity described or depicted; and

(b) has no moral, artistic or other value.

extreme crime and violence material, in relation to a computer game, means ~~material that is~~ crime and violence material ~~in relation to a computer game~~ where, without justification, the impact of the material is extreme because:

(a) the material is more detailed; or

(b) the material is realistic rather than stylised; or

(c) the game is highly interactive; or

(d) the gameplay links incentives or rewards to high impact elements of the game; or

(e) for any other reason.

extreme crime and violence material, in relation to a publication, means ~~material that is~~ crime and violence material ~~in relation to a publication~~ where, without justification, the impact of the material is extreme because of the emphasis, tone, frequency, context and detail of the relevant elements of the publication and other factors that heighten impact.

extreme crime and violence material, ~~in relation to material~~ that is not a computer game or a publication, means crime and violence material where, without justification, the impact of the material is extreme because:

(a) the material is more detailed; or

(b) the material is highly interactive; or

(c) the relevant depictions in the material are realistic, prolonged or repeated; or

(d) for any other reason.

general purpose DIS means a designated internet service that:

(a) is a website or application that:

(i) primarily provides information for business, commerce, charitable, professional, health, reporting news, scientific, educational, academic research, government, public service, emergency, or counselling and support service purposes;

(ii) enables transactions related to the matters in subparagraph (i); or

(b) is a web browser; and

(c) cannot be characterised as a different category of designated internet service under this industry standard.

high impact DIS means a designated internet service that:

(a) has the sole or predominant purpose of enabling end-users to access high impact materials; and

(b) makes available high impact material that has been posted by end-users;

and includes a service that is taken to be a high impact DIS because of subsection 13(2).

Note 1: This category would, for example, include websites or applications such as pornography websites and 'gore' or 'shock sites' that contain sexually explicit and/or graphically violent end-user generated content that qualifies as high impact material.

Note 2: Under paragraph 13(2)(a), a high impact DIS may also be taken to be a high impact generative AI DIS.

high impact generative AI DIS means a designated internet service:

(a) that uses machine learning models to enable an end-user to produce material; and

(b) for which it is reasonably foreseeable that the service could be used to generate synthetic high impact material.

and includes a service that is taken to be a high impact generative AI DIS because of subsection 13(2).

Note 1: This category would, for example, include services with generative artificial intelligence functionality to produce high impact material including completely new material and new material that has been created from editing existing material (for example – deepfake child sexual exploitation material).

Note 2: See note to definition of ***machine learning model platform service*** for example of an exclusion from this category.

Note 3: A high impact generative AI DIS may also be taken to be:

(a) a high impact DIS—see paragraph 13(2)(a); or

(b) a classified DIS—see paragraph 13(2)(b).

high impact materials, in relation to a high impact DIS, are materials which are:

(a) films or computer games which have been classified $R18^+$, $X18^+$ or RC in accordance with the Classification Act, or if classified would likely be classified as $R18^+$, $X18^+$ or RC; or

(b) publications which have been classified Category 1 Restricted, Category 2 Restricted, or RC in accordance with the Classification Act, or if classified would likely be classified Category 1 Restricted, Category 2 Restricted, or RC.

high impact materials, in relation to a high impact generative AI DIS are materials which are:

(a) films or computer games which have been classified $X18^+$ or RC in accordance with the Classification Act, or if classified would likely be classified as $X18^+$ or RC; or

(b) publications which have been classified Category 2 Restricted or RC in accordance with the Classification Act, or if classified would likely be classified Category 2 Restricted or RC.

industry code has the meaning given in section 132 of the Act.

justification: see subsection (4).

known child sexual abuse material means material that:

(a) is or includes images (either still images or video images); and

(b) has been verified as child sexual abuse material by a governmental (including multi-lateral) or non-governmental organisation:

(i) the functions of which are or include combating child sexual abuse or child sexual exploitation; and

(ii) in the case of a non-governmental organisation—that is generally recognised as expert or authoritative in that context; and

(c) is recorded on a database that:

(i) is managed by an organisation of a kind described in paragraph (b); and

(ii) is made available to government agencies, enforcement authorities and providers of designated internet services for the purpose of their using technological means to detect or manage child sexual abuse material on designated internet services.

Example: An example of a database referred to in paragraph (c) is the database managed by the National Center for Missing & Exploited Children.

known pro-terror material means material that has been verified as pro-terror material.

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

That List was accessible, on the registration of this industry standard, at https://www.un.org/securitycouncil/content/un-sc-consolidated-list.

Note 2: Material may, for example, be verified as a result of a decision of the Classification Board. Material may also be verified by using tools provided by independent organisations that are recognised as having expertise in counter-terrorism. Examples of these organisations include Tech against Terrorism and the Global Internet Forum to Counter Terrorism.

machine learning model platform service means a designated internet service with the predominant functionality of making available one or more machine learning models and making such models available for download.

Note: A machine learning model platform service which includes functionality to enable end-users to use a hosted model to generate synthetic high impact material is not considered a high impact generative AI DIS.

offensive: see subsection (5).

pre-assessed classified DIS means a classified DIS that meets the requirements of subsection (3).

pre-assessed tier 3 designated internet service means each of the following:

(a) a pre-assessed classified DIS; and

(b) a pre-assessed general purpose DIS.

pre-assessed general purpose DIS means a general purpose DIS that meets the requirements of subsection (3).

pro-terror material means:

(a) material that:

Commented [A12]: Refer to our and DIGI submissions.

Commented [A13]: Refer to our and DIGI submissions. Also, this definition moves the absolute standard of inclusion on the UNSC list (in RES Code) into a note and as a 'may'. This produces uncertainty re 'verification'. For example, in the current Gaza conflict, there may be material that is 'verified' by Arabic states as pro-terror that would not meet that definition. Similarly, not all material 'verified' by Israel may meet that definition.

Commented [A14]: Refer to our submission.

In order to address provider control (if AI/ML is indeed to be included into the standard, also refer to our submission), we suggest that a form of the Note from s23(3) which applies to many of the "new" AI services covered by the DIS, should appear in a standalone section in the standard and not be relegated to a Note. Suggested drafting (using the Note in 23(3)) is below.
"A requirement to put in place systems, processes, and technologies to minimise risk that class 1A material or class 1B material will be accessed or generated by, or distributed to, end-users in Australia using the service, or will be stored on the service, should take account of the fact that not all high impact generative AI DIS providers, enterprise DIS or machine learning model platform services will always have sufficient visibility and control of their models – if a provider lacks such visibility or control of certain aspects such that it cannot deploy all mitigations, it can rely on other systems, processes and technologies which are available."
In addition, MLMPS should have the same obligations as enterprise DIS and/or third party managed hosting as they have the same amount of control over as these providers.

Commented [A15]: Given the definition of machine learning model platform service (MLMPS) does not overlap with high impact gen AI DIS, it is unclear what is intended by this note.
We are also unclear about the significance of "a hosted model" in this context. This sentence implies that a MLMPS can be a service that uses hosted models, even though the definition above relates to making such models available for download (not hosted). This introduces unnecessary complexity and is inconsistent with the discussion paper accompanying the draft standards.

(i) directly or indirectly counsels, promotes, encourages or urges the doing of a terrorist act; or

(ii) directly or indirectly provides instruction on the doing of a terrorist act; or

(iii) directly praises the doing of a terrorist act in circumstances where there is a substantial risk that the praise might have the effect of a leading a person (regardless of the person's age or any mental impairment that the person might suffer) to engage in a terrorist act; or

(b) material that is known pro-terror material.

However, material accessible using a designated internet service is not pro-terror material if its availability on the service can reasonably be taken to be part of public discussion, public debate, entertainment or satire.

provide a designated internet service includes make the service available.

Commented [A16]: This definition is problematic in that many DIS make a service available but have no or limited control over the technical features of the service. The drafting of the compliance measures, however, references 'provide(r)' and, therefore, imposes requirements that cannot be complied with by the entity making a service available.

RC means the "Refused Classification" classification under the National Classification Code.

risk assessment means an assessment of a kind required by subsection 8(1).

risk profile, for a designated internet service, means the risk profile of the service worked out under subsection 8(8).

store: material is ***stored on a designated internet service*** if it is:

(a) in storage used for the service; or

(b) accessible through or using the service.

terms of use, for a designated internet service, means the provisions of the agreement under which the service is provided and includes anything that may reasonably be regarded as the equivalent of terms of use.

terrorist act has the meaning given by section 100.1(1) of the *Criminal Code* (no matter where the action occurs, the threat of action is made or the action, if carried out, would occur).

Tier 1 designated internet service means:

(a) a designated internet service that is determined in accordance with section 8 to have a Tier 1 risk profile;

(b) a high impact DIS; and

(c) a designated internet service that is determined in accordance with section 8(9) to have a Tier 1 risk profile.

Tier 2 designated internet service means a designated internet service that is determined in accordance with section 8 to have a Tier 2 risk profile.

Tier 3 designated internet service means:

(a) a designated internet service that is determined in accordance with section 8 to have a Tier 3 risk profile; and

(b) a pre-assessed Tier 3 designated internet service.

violence means an act of violence or an obvious threat of an act of violence.

Requirements for pre-assessment

(3) The requirements for a classified DIS or a general purpose DIS to be pre-assessed are that:

(a) in respect of posting or sharing of material—the relevant service

(i) does not enable end-users in Australia to post material to the service; or

(ii) 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; or

(iii) 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; and

(b) in respect of chat or messaging functionality—the relevant service:

(i) does not offer a chat or messaging function; or

(ii) offers a chat or messaging function but the chat or messaging function is limited to private messages or chats between the service and end-users in Australia for a business, informational, or government service or support purpose.

Justification

Commented [A17]: Refer to our and DIGI submissions in relation to context and Classification Scheme.

(4) For this industry standard, in determining whether material is without justification, the matters to be taken into account include:

(a) the standards of morality, decency and propriety generally accepted by reasonable adults; and

(b) the literary, artistic or educational merit (if any) of the material; and

(c) the general character of the material, including whether it is of a medical, legal or scientific character; and

(d) the persons or class of persons to or amongst whom it is published or is intended or likely to be published.

Offensive material

(5) The question whether material is offensive for the purposes of this industry standard is to be determined in accordance with the Act, including section 8 of the Act.

7 Technical feasibility

Commented [A18]: Refer to our and DIGI submissions.

In considering whether it is or is not technically feasible for the provider of a designated internet service to take a particular action, the matters to be taken into account include:

(a) the expected financial cost to the provider of taking the action; and

(b) whether it is reasonable to expect the provider to incur that cost, having regard to the extent of the risk to the online safety of end-users in Australia of not taking the action.

Part 3—Risk assessments and risk profiles

8 Requirement to carry out risk assessments and determine risk profiles of designated internet services

Risk assessments to be carried out

(1) The provider of a designated internet service must, at the times required by and in accordance with this Part, carry out an assessment of the risk that class 1A material or class 1B material:

(a) managed will be generated or accessed by, or distributed by or to, end-users in Australia using the service; and

(b) will be stored on the service.

Note: See also section 34.

Timing of risk assessments

(2) If the provider of the service was providing the service before the commencement of this industry standard, the risk assessment must be carried out as soon as practicable after, but no later than 6 months after, the commencement of this industry standard.

(3) Subsection (2) does not apply if a risk assessment that met the requirements of this Part had been carried out in respect of the service within 6 months before the commencement of this industry standard.

(4) A person must not start to provide a designated internet service to an end-user in Australia unless a risk assessment of the service has been carried out in accordance with this Part prior to the service being made available in Australia. ~~within 6 months before the person started to provide the service~~.

(5) The provider of a designated internet service must not make a material change to the service unless:

(a) a risk assessment of the service, as proposed to be changed, has been carried out in accordance with this Part; or

(b) the change will not materially increase the risk of class 1A material or class 1B material being accessed or generated by, or distributed to, end-users in Australia using the service, or being stored on the service.

Certain services exempt from risk assessment requirements

(6) Subsections (1) and (4) do not apply to any of the following:

(a) a pre-assessed general purpose DIS;

(b) a pre-assessed classified DIS;

(c) an end-user managed hosting service;

(d) an enterprise DIS;

(e) a high impact DIS;

(e) a high impact generative AI DIS;

Commented [A19]: Refer to our submission re risk mitigation.

Commented [A20]: Refer to our submission.

Commented [A21]: Refer to DIDI submission for a discussion of the test for changes.
Also, adding 'material' aligns with the reporting obligations further below.

(f) a machine learning model platform service;

(g) a designated internet service that is determined under subsection (9) to be a Tier 1 designated internet service.

Note: However, subsection (1) applies to the provider of a designated internet service mentioned in this subsection if it makes a material change to the service and this change would increase the risk of class 1A material or class 1B material being accessed or generated by, or distributed to end-users in Australia using the service, or being stored on the service ~~the service is materially changed~~.

Commented [A22]: s8(5) provides that a provider of a designated internet service mustn't make a material change unless either it has carried out a risk assessment OR there isn't an increase in the risk of class 1A or 1B material being access etc. What if one of the s8(6) exempt services does make a material change, but it doesn't raise the risk of class 1A or 1B material? Why would it then need to take a risk assessment? I've suggested wording to make the note under s8(6) consistent with s8(5)

Risk profiles of designated internet services

(7) The provider of a designated internet service that conducts a risk assessment of the service must, on completion of the assessment, determine, in accordance with subsection (8), what the risk profile of the service is.

(8) The risk profile of a designated internet service is worked out as follows:

Item	If the risk that class 1A material or class 1B material will be accessed or generated by, or distributed to, end-users in Australia using the service, or will be stored on the service, is…	the risk profile of the service is …
1	High	Tier 1
2	Moderate	Tier 2
3	Low	Tier 3

Note: Some designated internet services have a pre-assessed risk profile for purposes of this industry standard. For example, a high impact DIS is pre-assessed as having a Tier 1 risk profile, and a pre-assessed classified DIS and pre-assessed general purpose DIS is each pre-assessed as having a Tier 3 risk profile.

(9) However, the provider of a designated internet service may, at any time, without having conducted a risk assessment, determine that the risk profile of the service is Tier 1.

Note: See also section 34.

9 Methodology, risk factors and indicators to be used for risk assessments and risk profile determinations

Requirement for plan and methodology

(1) If the provider is required by this Part to carry out a risk assessment for a service, the provider must formulate in writing a plan, and a methodology, for carrying out the assessment that ensure that the risks mentioned in subsection 8(1) in relation to the service are accurately assessed.

(2) The provider must ensure that the risk assessment is carried out in accordance with the plan and methodology.

(3) The provider must ensure that a risk assessment is carried out by persons with the relevant skills, experience and expertise.

Forward-looking analyses of likely changes

(4) As part of a risk assessment carried out as required by this Part, the provider must undertake a forward-looking analysis of:

(a) likely changes to the internal and external environment in which the service operates or will operate, including likely changes in the functionality of, or the scale of, the service; and

(b) the impact of those changes on the ability of the service to meet the object of this industry standard.

Note: For the object of this industry standard see section 4.

Matters to be taken into account

(5) Without limiting subsection (1), the methodology for the conduct of a risk assessment must specify the principal matters to be taken into account in assessing relevant risks, which must include the following, so far as they are relevant to the service:

(a) the predominant purpose ~~functionality~~ of the service;

(b) the manner in which material is created or contributed to in connection with the service;

(c) the functionality of the service to enable end-users in Australia to post or share material;

(d) whether the service includes chat, messaging or other communications functionality;

(e) the extent to which material posted on or distributed using the service will be available to end-users of the service in Australia;

(f) the terms of use for the service;

(g) the terms of arrangements under which the provider acquires content to be made available on the service;

(h) the ages of end-users and likely end-users of the service;

(i) the outcomes of the analysis conducted as required by subsection (4);

(j) safety by design principles incorporated or relied upon during the design or operation of the service, including guidance and tools published or made available by a government agency or a foreign or international body;

(k) the risk to the online safety of end-users in Australia in relation to synthetic material generated by artificial intelligence.

Note 1: Arrangements referred to in paragraph (g) may include provisions that, if complied with, will reduce the risk that class 1A material and class 1B material will be made available through the service.

Note 2: Examples of agencies mentioned in paragraph (j) are the Commissioner or the Digital Trust & Safety Partnership Safe Framework.

Commented [A23]: Refer to our submission.

10 Documenting risk assessments and risk profiles

(1) As soon as practicable after determining the risk profile of a designated internet service, the provider of the service must record in writing:

(a) details of the determination; and

(b) details of the conduct of any related risk assessment;

sufficient to demonstrate that they were made or carried out in accordance with this Part.

(2) The record must include the reasons for the results of the assessment and the determination of the risk profile.

Note: See also section 34.

Part 4—Online safety compliance measures

Division 1—Preliminary

11 This Part not exhaustive

This Part does not prevent the provider of a designated internet service from taking measures, in addition to and not inconsistent with those required by this Part, to improve and promote online safety for Australians.

12 What is appropriate action?

(1) In determining whether action taken or proposed in relation to a designated internet service as required by this industry standard is appropriate, the matters to be taken into account include:

(a) the extent to which the action achieves the object of this industry standard in relation to the service; and

(b) if the action relates to a breach of applicable terms of use of a designated internet service, or community standards, in relation to class 1A material or class 1B material:

(i) the nature of the material and the extent to which the breach is inconsistent with online safety for end-users in Australia; and

(ii) the extent to which the action will or may reasonably be expected to reduce or manage the risk that the service will be used to solicit, generate, access, distribute or store class 1A material or class 1B material; and

(iii) whether the proposed action is proportionate to the level of risk to online safety for end-users in Australia from the material being accessible through the service.

Note: For the object of this industry standard see section 4.

Commented [A24]: Refer to DIGI submission.

13 Index of requirements for designated internet services

(1) The following table sets out the provisions of this Part applicable to providers of designated internet services.

Item	For this kind of designated internet service …	the applicable provisions of this Part are…
1	all designated internet services	sections 34, 40 and 41
2	Tier 1 designated internet service	(a) the provisions listed in item 1 (b) sections 14, 15, 17, 19, 20, 21, 22, 25, 26, 27, 28, 29, 30, 32, 33, 35, 36 and 37 (c) subsections 16(2) and (3) (d) subsections 18(2) and (3) (e) subsection 23(2)

Commented [A25]: It is unclear how these three compliance obligations will apply to pre-assessed services (e.g. pre-assessed classified DIS) given they are not required to complete a risk assessment, have complaints mechanisms, or do anything specific to comply with the standard. For example, s8(6) excludes pre-assessed general purpose DIS and pre-assessed classified DIS from risk assessment requirements, yet s13(1) provides that compliance obligation 34 (providing risk assessments to eSafety) is required for all DIS.
The best way to address this might be to create a new section for compliance obligations for pre-assessed Classified and General Purpose DIS. Some DIS will have no obligations other than where they make a material change to the service, and that is not acknowledged in the standard.

Item	For this kind of designated internet service …	the applicable provisions of this Part are…
		(f) subsections 24(1) to 4(a), and 24(5) to 24(6) (g) subsections 38(2) to (6), and 38(8)
3	Tier 2 designated internet service	(a) the provisions listed in item 1 (b) sections 14, 17, 19, 20, 28, 29, 31 and 32 (c) subsections 16(2) and (3) (d) subsections 18(2) and (3) (e) subsection 25(2) (f) subsections 38(2) to 38(6)
4	Tier 3 designated internet service	the provisions listed in item 1
5	end-user managed hosting service	(a) the provisions listed in item 1 (b) sections 14, 15, 17, 19, 20, 22, 26, 27, 28, 29, 31, 32, 35, 37 and 38 (c) subsections 16(2), (4), (5) and (6) (d) subsections 18(2) and (4) (e) subsections 21(2) to (8) (f) subsection 23(2) (g) subsection 25(2) (h) subsections 38(2) to (6), (8) and (9)
6	enterprise DIS	(a) the provisions listed in item 1 (b) section 14 (c) subsections 23(2) and (4) (d) subsections 38(2) to (6)
7	high impact generative AI DIS	(a) the provisions listed in item 1 (b) sections 14, 15, 17, 19, 20, 22, 24, 26, 27, 28, 29, 31, 32, 35 and 37 (c) subsections 16(2), (4) and (5) (d) subsections 18(2) and (4) (e) subsections 21(2) to (7) (f) subsections 23(2) and (3) (g) subsection 25(2) (h) subsections 38(2) to 38(6), 38(8) and 38(9)
8	machine learning model platform service	(a) the provisions listed in item 1 (b) sections 14, 15, 17, 20, 26, 27, 28, 29, 31, 37 and 38 (c) subsection 16(2) (d) subsection 23(2) (e) subsections 24(1) to (3), 24(4)(b) and 24(7) (f) subsection 25(2) (g) subsections 38(2) to (7)

Note 1: Subsection 24(4)(b) does not apply to a Tier 1 designated internet service.

Note 2: Subsection 24(4)(a) does not apply to a machine learning model platform service.

(2) Where a designated internet service meets the definition of more than one kind of designated internet service under this industry standard, for the purposes of this industry standard:

(a) if the service meets the definition of a high impact DIS and a high impact generative AI DIS—the service is taken to be a service of each of those kinds;

(b) if the service meets the definition of a classified DIS and a high impact generative AI DIS—the service will be taken to be a service of each kind;

(c) if the service meets the definition of an enterprise DIS and another type of designated internet service ~~an end-user managed hosting service~~—the service will be:

(i) if made available primarily to enterprise customers—taken to be an enterprise DIS; and

(ii) if primarily made available by the provider directly to end-users in Australia—taken to be the other type of designated internet service ~~an end-user managed hosting service~~; and

(d) if the service meets the definitions of 2 or more other designated internet services—the service will be taken to be the kind of designated internet service that is most closely aligned with the service's predominant purpose ~~functionality~~.

Note 1: For paragraphs (a) and (b), this means the provider of the service must ensure the service meets the minimum compliance measures that are applicable to each kind of service.

Note 2: For paragraph (c), this means that the provider of the service must ensure the services meets the minimum compliance measures applicable to:

(a) an enterprise DIS (when the service is primarily being provided to enterprise customers); and

(b) an end-user managed hosting service (when the service is being primarily provided directly to end-users).

Commented [A26]: Throughout the standard note the inconsistent use of 'make available' and 'provide' but also refer to our comments in relation to the definition of 'provide'.

Commented [A27]: The functionality argument does not work well for an enterprise DIS (i.e. the logic in Note 2 applies in all cases where an enterprise DIS is also being offered directly to end-users as another type of DIS). We suggest amendments to (c) to address this. We do not see a reason why (c) should only be limited to enterprise DIS vs. end user managed hosting DIS.

Division 2—Minimum compliance measures—general

14 Terms of use

(1) This section applies to the following services:
  (a) a Tier 1 designated internet service;
  (b) a Tier 2 designated internet service;
  (c) an end-user managed hosting service;
  (d) an enterprise DIS;
  (e) a high impact generative AI DIS; and
  (f) a machine learning model platform service.

Provisions to be included in terms of use

(2) The provider of a service must include provisions in the terms of use for the service that:
  (a) impose an obligation on the account holder of the service to ensure that the service is not used, whether by the account holder or an end-user in Australia, 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
  (b) give rights for the provider to do any of the following if the service is used to solicit, access, generate, distribute or store (as applicable) class 1A material or class 1B material:
    (i) suspend the provision of the service to a specified end-user of the service for a specified period;
    (ii) impose specified restrictions on the use of the service by a specified end-user of the service for a specified period;
    (iii) terminate the agreement for the provision of the service.

Enforcement of terms of use

(3) If the provider of a service becomes aware of a breach of the obligation mentioned in subsection (2)(a), the provider must enforce its contractual rights in respect of the breach in an appropriate way that is proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach.

(4) In proceedings in respect of a contravention of subsection (3), the provider bears the evidential burden of establishing:
  (a) the action it took to enforce the rights; and
  (b) that the action that it took was appropriate and proportionate, as referred to in subsection (2).

Note: For appropriate action see also section 12.

Commented [A28]: Given the similarity between third party hosting and enterprise DIS and MLMPS we suggest adding the same qualifiers from the hosting code here. Specifically:

"For the purpose of this measure, providers may satisfy this measure in different ways and by making use of different language. Providers may consider that existing language in polices and/or contractual terms satisfies this requirement.
Guidance:
Providers have flexibility to design terms and policies to allow appropriate and proportionate responses to potential breaches on a case-by-case basis. Providers have the ability to exercise discretion to enforce terms and policies in accordance with the specific circumstances of each potential breach."

15 Notification of child sexual exploitation material and pro-terror material

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) an end-user managed hosting service;
(c) a high impact generative AI DIS;
(d) a machine learning model platform service.

(2) If the provider of a service:
(a) identifies child sexual exploitation material, or pro-terror material, on the service; and
(b) believes in good faith that the material affords evidence of a serious and immediate threat to the life or physical safety of a person in Australia;
the provider must, as soon as practicable, report the matter to an enforcement authority, or otherwise as required by law.

(3) If the provider of a service:
(a) identifies child sexual abuse ~~exploitation~~ material on the service; ~~and~~
(b) believes in good faith that the material is reasonably likely to be ~~not known~~ child sexual abuse ~~exploitation~~ material; and
(c) believes in good faith that the material is not known sexual abuse material;
the provider must, as soon as practicable, notify an organisation of a kind referred to in paragraph (b) of the definition of known child sexual abuse ~~exploitation~~ material in subsection 6(1).

(4) If the provider of a service:
(a) identifies pro-terror material on the service; and
(b) believes in good faith that the material is not known pro-terror material;
the provider must, as soon as practicable, notify an organisation that verifies material as pro-terror material.

Note: See the definition of ***pro-terror material*** in subsection 6(1).

(5) Subsections (2), (3) and (4) are in addition to any other applicable law.

16 Systems and processes for responding to breaches of terms of use or community standards: class 1A material

(1) This section applies to the following services:
(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS;
(e) a machine learning model platform service.

Commented [A29]: Refer to DIGI submission for a discussion of available databases, authorities, context etc. required for the detection of such material and notification to authorities.

Commented [A30]: Refer to our submission.
An MLMPS does not control a model once it's downloaded so it is not possible to meet this obligation. The Note from MCM 23(3) (extracted below) applies to many of the "new" AI services covered by the DIS and should have it's own standalone section in the standard and not be relegated to a Note (see proposed drafting under MLMPS definition)

"Note: A requirement to put in place systems, processes, and technologies to disrupt and deter the production of CSEM should take account of the fact that not all high impact generative AI DIS providers will always have sufficient visibility and control of their models – if a provider lacks such visibility or control of certain aspects such that it cannot deploy all mitigations, it can rely on other systems, processes and technologies which are available."

Commented [A31]: Child sexual exploitation material (being the broadest category) is already dealt with in (2) so this section should be limited to CSAM. Depictions of child-like characters (e.g. erotic manga) should not be reported to NCMEC but under the proposed eSafety drafting, this would be a requirement. This requirement should really only capture actual abuse material.

Minimum requirements—generally

(2) The provider of a service must implement systems and processes that ensure that, if the provider becomes aware that:

(a) there is or has been a breach, of an obligation under the terms of use for the service in respect of class 1A material, including a breach of an obligation to comply with acceptable use policies; or

(b) there is or has been a breach, involving the service, of community standards in respect of class 1A material;

the provider takes appropriate action to ensure that:

(c) the breach, if it is continuing, ceases; and

(b) the risk of further such breaches is minimised.

Further minimum requirements—Tier 1 or Tier 2 designated internet services

(3) Without limiting subsection (2), the systems and processes implemented by a provider of a Tier 1 or Tier 2 designated internet service must include ones under which the provider:

(a) reviews reports by end-users of the service in Australia that class 1A materials are accessible using the service; and

(b) appropriately prioritises those reports and, if necessary, escalates them to senior management personnel of the provider for action.

Note: For paragraph (a), reports include the reports referred to in section 29.

Further minimum requirements —end-user managed hosting services and high impact generative AI DIS

(4) Without limiting subsection (2), a provider of an end-user managed hosting service or high impact generative AI DIS must establish and implement standard operating procedures that:

(a) require the provider to engage with reports of class 1A material received from end-users to help determine whether the terms of use of the service (including acceptable use policies), or community standards, prohibiting class 1A material on the service have been breached; and

(b) enable the provider to take appropriate action to assess and respond to those breaches.

~~(5) A provider of an end-user managed hosting service or high impact generative AI DIS must implement the standard operating procedures established as required by subsection (4).~~

Further minimum requirements —end-user managed hosting services

(5~~6~~) Without limiting subsections (2) to (5), the provider of an end-user managed hosting service must implement practices and procedures that are appropriate to minimise the likelihood that class 1A material is accessible by end-users of the service. This includes ensuring that terms of use for the service that prohibit the storage or hosting of class 1A material on the service are in place with:

(a) for an enterprise DIS—the account-holders; and
(b) in other cases—the end-users.

17 Responding to breaches of terms of use or community standards—CSEM and pro-terror material

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS;
(e) a machine learning model platform service.

Commented [A32]: Remove MLMPS because there is no ability to remove content and there is no relationship to the end user to terminate their account.

(2) If the provider of a service becomes aware that:
(a) there is or has been a breach of an obligation under the terms of use for the service in respect of CSEM or pro-terror material, including a breach of an obligation to comply with acceptable use policies; or
(b) there is or has been a breach, involving the service, of community standards in respect of CSEM or pro-terror material;

the provider must:
(c) remove instances of CSEM and pro-terror materials identified by the provider on the service as soon as reasonably practicable unless otherwise required to deal with unlawful CSEM and pro-terror materials by an enforcement authority;
(d) terminate an end-user's account as soon as reasonably practicable if the end-user:
(i) is distributing CSEM or pro-terror materials to end-users with the intention to cause harm;
(ii) is known to be an Australian child using the account; or
(iii) has repeatedly breached terms and conditions, community standards or acceptable use policies prohibiting CSEM and pro-terror materials on the service.

(3) Without limiting subsection (2), the provider must take appropriate action to ensure that:
(a) the service no longer permits access to or distribution of the material; and
(b) the breach, if it is continuing, ceases; and
(c) the risk of further such breaches is minimised; and
(d) end-users who repeatedly breach terms of use, community standards or acceptable use policies prohibiting CSEM or pro-terror material and who have had their user accounts terminated, do not acquire new accounts.

(4) Without limiting what is appropriate action, appropriate action may include the provider exercising, in a way that is proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach, any of the providers contractual rights under the terms of use for the service in relation to the breach.

Commented [A33]: Unclear why we need this when this is already covered in s14?

Note: For the contractual rights required to be included in terms of use see paragraph 14(1)(b).

18 Responding to breaches of terms of use or community standards—class 1B material

(1) This section applies to the following:
  (a) a Tier 1 designated internet service;
  (b) a Tier 2 designated internet service;
  (c) an end-user managed hosting service;
  (d) a high impact generative AI D IS.

Minimum requirements—generally

(2) The provider of a service must implement systems and processes that ensure that, if the provider becomes aware that:
  (a) there is or has been a breach, in Australia, of an obligation under for the terms of use for the service in respect of class 1B material, including a breach of an obligation to comply with acceptable use policies; or
  (b) there is or has been a breach, in Australia, involving the service, of community standards in respect of class 1B material;

the provider takes appropriate action to ensure that:
  (c) the breach, if it is continuing, ceases; and
  (d) the risk of further such breaches is minimised.

Further minimum requirements —Tier 1 or Tier 2 designated internet services

(3) Without limiting subsection (2) the systems and processes implemented by a provider of a Tier 1 or Tier 2 designated internet service must:
  (a) include ones under which the provider:
    (i) reviews reports by end-users of the service in Australia that class 1B materials are accessible using the service; and
    (ii) appropriately prioritises those reports and, if necessary, escalates them to senior management personnel of the provider for action; and
  (b) include operational guidance to provider personnel, including actions to be take and time limits to be observed, in performing the provider's duties under this section.

Further minimum requirements for end-user managed hosting services and high impact generative AI DIS

(4) Without limiting subsection (2), the provider of an end-user managed hosting service or high impact generative AI DIS must implement standard operating procedures that:
  (a) require the provider to engage with reports of class 1B material received from end-users to help determine whether the provider's terms and conditions, community standards, and/or acceptable use policies relating to class 1B materials on the service have potentially been breached and

(b) enable the provider to take appropriate action to assess and respond to potential breaches of terms and conditions, community standards, and/or acceptable use policies prohibiting class 1B material.

19 Action in response to breaches of policies relating to extreme crime and violence material and class 1B material

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS.

(2) If the provider of a service becomes aware that:
(a) there is or has been a breach, in Australia, of an obligation under the terms of use for the service in respect of extreme crime and violence material or class 1B material, including a breach of an obligation to comply with acceptable use policies; or
(b) there is or has been a breach, in Australia, involving the service, of community standards in respect of extreme crime and violence material or class 1B material;
the provider must take appropriate action to respond to the breach.

20 Resourcing trust and safety functions

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS;
(e) a machine learning model platform service.

(2) The provider of a service must have and implement, in respect of the service, management, supervision and internal reporting arrangements to ensure that at all times the provider:
(a) complies with the requirements of this industry standard; and
(b) can otherwise effectively supervise the online safety of the service.

Note: These arrangements may include duties and responsibilities for personnel, and systems, processes or~~and~~ technologies.

(3) The provider of a service must have, or have access to, sufficient personnel who have the skills, experience and qualifications needed to ensure that the provider complies with the requirements of this industry standard at all times.

21 Detecting and removing known CSAM

(1) This section applies to the following services:

(a) a Tier 1 designated internet service;
(b) an end-user managed hosting service;
(c) a high impact generative AI DIS.

Minimum requirements—generally

(2) The provider of a service must implement systems, processes and technologies that detect and identify instances of known CSAM that:
(a) is stored on the service; or
(b) is accessible by an end-user in Australia using the service; or
(c) is being or has been accessed or distributed in Australia using the service.

Note 1: Such systems, processes and technologies include for example using hashing, machine learning, artificial intelligence systems that scan for known CSAM.

Note 2: For a high impact generative AI DIS, compliance with this subsection may require the provider to assess whether inputs into the service contain known child sexual abuse material.

Commented [A34]: Note 2, "assessing inputs" may happen before any of Section 21(2)(a-c) happen. And the content is not "removed" because it never enters the provider's control, it's simply "denied." Therefore, by definition Note 2 is not covered by Section 21(2).

(3) Subsection (2) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(4) The provider of a service must implement systems, processes and technologies that remove known CSAM from the service as soon as practicable after it is detected and identified.

(5) Subsection (4) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(6) If it is not technically feasible for the provider to implement a particular system, process or technology for the purposes of:
(a) detecting and identifying known CSAM as required by subsection (2); or
(b) removing known CSAM as required by subsection (4);
the provider must take appropriate alternative action.

Note: For appropriate action see section 12.

(7) This section does not affect the operation of section 23.

Note 1: For technical feasibility, see section 7.

Note 2: This section does not prevent a provider from complying with legal obligations to preserve evidence of offences.

Further minimum requirements —Tier 1 designated internet services and end-user managed hosting services

(8) Without limiting subsections (2) to (7), the provider of a Tier 1 designated internet service or an end-user managed hosting service must ensure the service uses systems, processes, and technologies that automatically detect and flag known CSAM.

Commented [A35]: This adds significant technical and financial burdens to SMEs and, so we believe, is an inappropriate catch all provision.

Further minimum requirements —Tier 1 designated internet services

(9) Without limiting subsection (2) to (8), the provider of Tier 1 designated internet service must ensure the service uses systems, processes, and technologies that:

(a) prevent end-users from distributing known CSAM; and
(b) identify phrases or words commonly linked to CSAM and linked activity to enable the provider to deter and reduce the incidence of such material and linked activity.

22 Detecting and removing known pro-terror material

Commented [A36]: Refer to our and DIGI submissions for feedback in relation to pro-terror material.

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) an end-user managed hosting service; and
(c) a high impact generative AI DIS.

(2) The provider of a service must implement systems, processes, and technologies that detect and identify known pro-terror material that:
(a) is stored on the service; or
(b) is accessible by an end-user in Australia using the service; or
(c) is being or has been generated, accessed or distributed in Australia using the service.

Note 1: Such systems, processes and technologies include for example using hashing, machine learning, artificial intelligence systems that scan for known pro-terror material.

Note 2: For a high impact generative AI DIS, compliance with this subsection may require the provider to assess whether inputs into the service contain known pro-terror material.

(3) Subsection (2) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(4) The provider of a service must implement systems, processes and technologies that remove instances of known pro-terror material from the service as soon as practicable after it is detected and identified.

(5) Subsection (4) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(6) If it is not technically feasible for the provider to implement a particular system, process or technology for the purposes of:
(a) detecting and identifying known pro-terror material as required by subsection (2); or
(b) removing known pro-terror material as required by subsection (4);
the provider must take appropriate alternative action.

Note: For appropriate action see section 12.

(7) This section does not affect the operation of section 23.

Note 1: For technical feasibility, see section 7.

Note 2: This section does not prevent a provider from complying with legal obligations to preserve evidence of offences.

23 Disrupting and deterring CSEM and pro-terror material

(1) This section applies to the following:

(a) a Tier 1 designated internet service;
(b) an end-user managed hosting service;
(b) an enterprise DIS;
(c) a high impact generative AI DIS;
(d) a machine learning model platform service.

Minimum requirements—generally

(2) The provider of a service must implement systems, processes and technologies that:
(a) effectively deter end-users of the service from using the service; and
(b) effectively disrupt attempts by end-users of the service to use the service;
to solicit, generate, access, distribute, store or otherwise make available CSAM and pro-terror material (including known CSAM and known pro-terror material).

Note: Examples of systems, processes and technologies include hashing, machine learning, artificial intelligence systems that scan for known CSAM and those that are designed to detect key words, behavioural signals and patterns associated with child sexual abuse material.

Further minimum requirements —high impact generative AI DIS

(3) Without limiting subsection (2), the provider of a high impact generative AI DIS must, at a minimum:
(a) implement systems, processes and technologies that prevent generative AI features from being used to generate outputs that contain CSEM and pro-terror material;
(b) regularly review and test models on the potential risk that a model is used to generate CSEM and pro-terror material;
(c) promptly following review and/or testing, adjust models and deploy mitigations with the aim of reducing the misuse and unintentional use of models to generate CSEM and pro-terror material;
(d) implement systems, processes and technologies that differentiate AI outputs generated by the model;
(e) ensure that end-users in Australia specifically seeking images of CSAM are presented with prominent messaging that outlines the potential risk and criminality of accessing CSAM; and
(f) ensure that material generated for end-users in Australia prompts using terms that have known associations to CSEM are accompanied by information or links to services that assist end-users in Australia to report CSEM to enforcement agencies and/or seek support; and
(g) ensure that the systems and processes implemented by the provider under subsection (2) are able to automatically detect and action CSAM in training data, user prompts, and outputs, with the aim of preventing this material from being generated, for example, using hashing, key word lists, classifiers, or other safety technologies.

Note: A requirement to put in place systems, processes, and technologies to disrupt and deter the production of CSEM should take account of the fact that not all high impact generative AI DIS providers will always have sufficient visibility and control of their models – if a provider lacks such visibility or control of certain aspects such that it

Commented [A37]: This does not work for enterprise DIS or MLMPS because neither has a relationship with the "end user of the service" which is what 23(2)(a) and (b) relate to. Both enterprise DIS and MLMPS have a contractual relationship with the "account holder" but operate more like a third-party hosting provider. Byond issues around control, it is simply not meaningful for e.g. providers of enterprise DIS such as websites designed for the ordering of commercial supplies by enterprise customers to implement measures to this effect.

Commented [A38]: See comment above - the MLMPS does not have a relationship with the "end-user" so can't comply with (2) (a) and (b). Delete MLMPS.

Commented [A39]: As per above, this doesnt make sense for enterprise DIS. Amend heading to only include the services this is relevant for (i.e. Tier 1, end-user managed hosting services, high impact generative AI DIS).

Commented [A40]: The Note mixes known tech (hashing, ML, and keywords) with "smart tech" that not everyone has access to. "Behavioral signals" and "patterns" assume the provider is tracking users over time, which may violate privacy laws.

Commented [A41]: Note many services do not provide services specifically to Australian end-users - they are global services with a global user base. Where it would be illegal under the laws of other countries to implement these measures, this needs to be addressed.

cannot deploy all mitigations, it can rely on other systems, processes and technologies which are available.

Note: For paragraph (d), systems, processes and technologies may include by embedding indicators of provenance into material generated by a model to enable differentiation.

~~*Further m*~~*Minimum requirements —enterprise DIS*

(4) ~~Without limiting subsection (2),~~ the provider of an enterprise DIS which provides pre-trained machine learning models for integration into a service deployed or to be deployed by an enterprise customer must, at a minimum:

(a) implement and use systems, processes and technologies that ~~automatically~~ detect and flag CSAM in training data; and

(b) take appropriate action to ensure the service cannot be used to generate CSAM based on, using or otherwise related to such CSAM.

Note: See note 2(b) to the definition of ***enterprise DIS***.

24 Development programs

(1) This section applies:

(a) to the following:

(i) a Tier 1 designated internet service;

(ii) a high impact generative AI DIS;

(iii) a machine learning model platform service;

but only where the average monthly number of active end-users of the service, in Australia, over the immediate previous calendar year was 1,000,000 or more; and

(b) to an end-user managed hosting service but only where the average monthly number of active end-users of the service, in Australia, over the immediate previous calendar year was 500,000 or more.

Commented [A42]: Refer to our submission.

(2) However:

(a) the provisions of this section, so far as they relate to pro-terror material, do not apply to a Tier 1 designated internet service predominantly used for making pornography available;

(b) paragraph 4(b) does not apply to a Tier 1 designated internet service; and

(c) paragraph 4(a) does not apply to a machine learning model platform service.

(3) The provider of the service must establish and implement, for the calendar year, a program of investment and development activities (***development program***) in respect of systems, processes and technologies.

Note: See also section 37.

(4) A development program must include:

(a) investments and activities designed to develop systems, processes and technologies that enhance the ability of the provider, or of other providers of designated internet services:

(i) to detect and identify child sexual abuse material or pro-terror material (including known child sexual abuse material and known pro-terror material) on the service; and

(ii) effectively to deter end-users of the service from using the service, and to disrupt attempts by end-users of the service to use the service, to generate, access, distribute or store child sexual abuse material or pro-terror material (including known child sexual abuse material and known pro-terror material); and

(b) arrangements for cooperating and collaborating with other organisations in activities of the kind referred to in paragraph (a) and to enhance online safety for Australians.

Commented [A43]: Can this be discharged by participation in the industry associations' annual fora (Codes) or what exactly is the expectation?

(5) A development program may include arrangements for the provider to make available to other providers of designated internet services, or organisations engaged in promoting online safety for Australians, systems, processes and technologies of a kind referred to in paragraph (4)(a) (including making them available without charge).

(6) Examples of investments and activities that may be part of a provider's development program for purposes of paragraph 4(a) include:

(a) procuring online safety systems and technologies for use in connection with the service, or enhancing online safety systems and technologies used in connection with the service; and

(b) conducting research into and development of online safety systems and technologies; and

(c) providing support, either financial or in kind, to organisations the functions of which are or include working to combat child sexual abuse, child sexual exploitation or terrorism.

Note: For paragraph (c), other organisations can include universities, the CSIRO, the WePROTECT Global Alliance, and the Global Internet Forum to Counter Terrorism (GIFCT).

(7) Examples of arrangements that may be part of a provider's development program for purposes of paragraph 4(b) include:

(a) joining industry organisations intended to address serious online harms; and

(b) sharing information on best practice approaches, that are relevant to the service; and

(c) working with the Commissioner to share information, intelligence, best practices and other information relevant to addressing categories of class 1A material or class 1B material that are relevant to the service; and

(d) collaborating with non-government or other organisations that facilitate the sharing of information, intelligence, best practices and other information relevant to addressing categories of class 1A or class 1B material that are relevant to the service.

25 Safety features and settings—class 1A material and class 1B material

(1) This section applies to the following services:

(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS; and
~~(e) a machine learning model platform service.~~

Commented [A44]: As above, MLMPS does not have control once models are downloaded and cannot meet this MCM. We suggest deletion.

Minimum requirements—generally

(2) The provider of a service must:
(a) carry out an assessment of the kinds of features and settings that could be incorporated into the service to minimise the risk that class 1A material and 1B material:
(i) will be accessed by, or distributed to, end-users in Australia using the service; or
(ii) will be stored on the service; and
(b) determine, on the basis of the assessment, the most appropriate and effective features and settings for the service; and
(c) ensure that the service at all times incorporates the features and settings so determined.

Further minimum requirements for a provider of a Tier 1 designated internet service

(3) Without limiting subsection (2), the provider of a Tier 1 designated internet service must:
(a) implement measures that ensure that material can only be posted to or distributed on the service by a registered account holder; and
(b) make clear in terms and conditions, community standards, and/or acceptable use policies that an Australian child is not permitted to hold an account on the service.

(4) The provider of the service must take appropriate action to:
(a) ensure that a child in Australia who is known by the Provider to be under the age of 18 does not become an end-user of the service; and
(b) stop access to the service by a child in Australia who is known by the provider to be under the age of 18.

26 Referral of unresolved complaints to the Commissioner

(1) This section applies to the following services:
(a) a Tier 1 designated internet service;
(b) an end-user managed hosting service;
(c) a high impact generative AI DIS; and
~~(d) a machine learning model platform service.~~

Commented [A45]: An MLMPS cannot resolve complaints given lack of control so we suggest it be removed from this section.

(2) The provider of a service must refer complaints from individuals (including end-users) concerning the provider's non-compliance with this industry standard to

the Commissioner where the complaint is not resolved within a reasonable period.

27 Responding and referring to the Commissioner

(1) This section applies to the following:
- (a) a Tier 1 designated internet service;
- (b) an end-user managed hosting service;
- (c) a high impact generative AI DIS;
- (d) a machine learning model platform service.

(2) The provider of a service must implement policies and procedures that ensure that:
- (a) it responds in a timely and appropriate manner to communications from the Commissioner about compliance with this industry standard; and
- (b) it refers unresolved complaints to the Commissioner in accordance with section 26.

28 Giving information about the Commissioner to end-users in Australia

(1) This section applies to the following:
- (a) a Tier 1 designated internet service;
- (b) a Tier 2 designated internet service;
- (c) an end-user managed hosting service;
- (d) a high impact generative AI DIS;
- ~~(e) a machine learning model platform service.~~

(2) The provider of a service must ensure that information:
- (a) describing the role and functions of the Commissioner; and
- (b) describing how to make a complaint to the Commissioner about the service; and
- (c) describing the mechanisms and process required by section 29 for the service;

is accessible to end-users of the service in Australia at all times through a dedicated location on the website for the service. The location must be "in service", that is, not on a separate webpage to the webpage for the service.

29 Mechanisms for end-users and account holders to report, and make complaints about, information on designated internet services

(1) This section applies to the following:
- (a) a Tier 1 designated internet service;
- (b) a Tier 2 designated internet service;
- (c) an end-user managed hosting service;
- (d) a high impact generative AI DIS;
- ~~(e) a machine learning model platform service.~~

Commented [A46]: Same comment as above re lack of control of MLMPS. MLMPS should have the same requirements as enterprise DIS.

Commented [A47]: This may make sense for complaints mechanisms but not for the other information required for this section.
It's also unclear what "in service" means in this context. The example "not on a separate webpage to the webpage for the service' assumes all DIS are websites which is not the case (e.g. an app can also be a DIS). This language is too precise/specific and not flexible enough for companies to enable compliance.

Commented [A48]: As above for MLMPS. MLMPS do not have a contractual relationship with end users of the service.

(2) The provider of a service must implement and make available a mechanism that:
  (a) enables end-users of the service in Australia to report, flag, or make a complaint about material accessible on the service on the basis that the material:
    (i) is class 1A material or class 1B material; and
    (ii) is in breach of an obligation under the terms of use for the service, including an obligation to comply with acceptable use policies; and
  (b) is easily accessible on or through the service, and easy to use; and
  (c) includes or is accompanied by clear instructions on how to make a report or complaint, and an overview of the reporting and complaints process.

Note: for paragraph (a) for a high impact generative AI DIS, ***material*** includes material generated (or capable of being generated) by the service.

(3) The provider of the service must ensure that the identity of a person who makes a report or a complaint using the mechanism under subsection (2) (the ***first end-user***) is not accessible, directly or indirectly, by any other end-user of the service without the express consent of the first end-user.

(4) The provider of the service must:
  (a) document its systems, processes and technologies dealing with how it responds to reports made under paragraph 29(2)(a); and
  (b) ensure that personnel responding to reports made under paragraph 29(2)(a) have appropriate training in and experience of the provider's policies and procedures for dealing with reports.

30 Action in response to end-user reports—Tier 1 designated internet services

(1) This section applies to a Tier 1 designated internet service.

(2) The provider of a service must:
  (a) take appropriate action to respond promptly to reports made by end-users under paragraph 29(2)(a); and
  (b) ensure that an end-user who makes a report concerning class 1A or class 1B materials:
    (i) is notified promptly of the outcome of the report; and
    (ii) is able to request a review by the provider of outcome under paragraph (i); and
    (iii) is notified promptly of the outcome of a review under paragraph (ii).

Note 1: A report includes a request for a review of the outcome of a report under paragraph (b).

Note 2: Without limiting section 12, appropriate action may include regular reviews of the effectiveness of the measures implemented by the service provider to ensure compliance with this section.

31 Action in response to end-user reports – other designated internet service providers

(1) This section applies to the following:
  (a) a Tier 2 designated internet service;

(b) an end-user managed hosting service;
(c) a high impact generative AI DIS;
~~(d) a machine learning model platform service.~~

(2) The provider of a service must take appropriate action to respond promptly to reports made by an end-user under paragraph 29(2)(a);

Commented [A49]: Same comment above re MLMPS and lack of control to receive or action reports from "end users".

32 Policies and terms of use to be published

(1) This section applies to the following:
(a) a Tier 1 designated internet service;
(b) a Tier 2 designated internet service;
(c) an end-user managed hosting service;
(d) a high impact generative AI DIS.

Note: For paragraph (d), for a high impact generative AI DIS, ***material*** includes material generated (or capable of being generated) by the service.

(2) The provider of a service must publish:
(a) the terms of use for the service, including provisions relating to acceptable use policies; and
(b) a statement setting out the community standards applicable to the service.

(3) The publication must be accessible on the website and application (if any) for the service.

(4) The publication must:
(a) be in plain English; and
(b) make it clear that class 1A material is not permitted on the service and describe the broad categories of material within class 1A material; and
(c) describe the broad categories of material within class 1B material and specify the extent to which that material is not permitted on the service, or is subject to specified restrictions.

Note: For paragraphs (b) and (c) for a high impact generative AI DIS, ***material*** includes material that is not permitted to be generated by the service.

33 Information on actions taken by the provider

(1) This section applies to a Tier 1 designated internet service.

(2) The provider of a service must publish, through a dedicated location on the website for the service, information on the:
(a) systems, processes and mechanisms implemented by the provider; and
(b) other actions taken, or to be taken, by the provider;
to reduce the risk of end-users accessing, generating or being exposed to class 1A material and class 1B material through the service.

Commented [A50]: Technologies? Consistency of terminology.

Division 3—Reporting requirements

34 Commissioner may require risk assessments and other information

(1) This section applies to all designated internet services to which this industry standard applies.

(2) The Commissioner may, by notice to the provider of a designated internet service, require the provider to give the Commissioner any of the following documents:

(a) the most recent risk profile determination for the service;

(b) the record, as required by section 10, of the most recent risk assessment for the service;

(c) the applicable risk methodology for the most recent risk assessment for the service;

(d) the provider's development program for a specified calendar year.

Note: For development programs see section 23.

(3) The provider must give the documents to the Commissioner within the period specified in the notice.

Note 1: See also section 39.

Note 2: A provider of a designated internet service that is not required to prepare certain documents under this standard, such as risk assessments, will not be required to give the Commissioner such documents.

35 Reports of technical feasibility of provisions of Division 2

(1) This section applies to the following:

(a) a Tier 1 designated internet service;

(b) an end-user managed hosting service;

(c) a high impact generative AI DIS.

(2) The Commissioner may, by notice to the provider of a designated internet service, require the provider to give the Commissioner a report that specifies the extent to which it is technically feasible for the provider to comply with a specified provision of Division 2.

(3) If the report discloses that it is not, or has not been, technically feasible for the provider to use a system, process or technology as required by subsection 21(2) or (4), the report must specify the alternative action required by subsection 21(6).

Note: Section 21 is about known child sexual abuse material.

(4) If the report discloses that it is not, or has not been, technically feasible for the provider to use a system, process or technology as required by subsection 22(2) or (4), the report must specify the alternative action required by subsection 22(6).

Note: Section 22 is about known pro-terror material.

(5) The report must provide justification for the conclusions in the report.

Commented [A51]: Refer to our submission regarding missing confidentiality provisions.

Commented [A52]: As indicated above, s 34 purports to apply to all DIS, including pre-assessed DIS which do not need to undertake risk assessments unless they make a material change to the service. Suggest adding the following additional note:
"A provider of a designated internet service that is not required to prepare certain documents under this standard, such as risk assessments, will not be required to give the Commissioner such documents."

Commented [A53]: We believe a minimum notice period of 2 months (or 60 days) should be specified.

(6) The notice may require the report to be in a specified form. The provider must comply with the requirement.

(7) A report may relate to 2 or more services.

(8) The provider must give the report to the Commissioner within the period specified in the notice.

Note: See also section 39.

Commented [A54]: We believe a minimum notice period of 2 months (or 60 days) should be specified.

36 Notifying new features of designated internet services

(1) This section applies to a Tier 1 designated internet service.

(2) If the provider of a service decides to add a new feature or function to the service, the provider must notify the Commissioner of the proposed change as soon as practicable after making the decision unless the provider considers, on reasonable grounds, that the proposed change will not ~~significantly~~ materially increase the risk that the service will be used to solicit, access, generate, distribute or store class 1A material or class 1B material.

Commented [A55]: Use materiality concept.

(3) If a new feature or function is added to a service, the provider of the service must notify the Commissioner of the change as soon as practicable unless the provider determines, on reasonable grounds, that the change has not ~~significantly~~ materially increased the risk that the service will be used to solicit, access, generate, distribute or store class 1A material or class 1B material.

37 Reports on outcomes of development programs

(1) This section applies to the following services:
- (a) a Tier 1 designated internet service;
- (b) an end-user managed hosting service;
- (c) a high impact generative AI DIS; and
- (d) a machine learning model platform service.

(2) The Commissioner may, by notice to the provider of a designated internet service to which section 24 applied in respect of a particular calendar year, require the provider to give the Commissioner a report that specifies:
- (a) the activities undertaken by the provider in respect of the calendar year to implement its development program; and
- (b) the outcomes of those activities in terms of enhancing online safety for end-users in Australia.

(2) The Commissioner may, by notice to the provider, require the report to be in a specified form. The provider must comply with the requirement.

(4) The provider must give the report to the Commissioner within the period specified in the notice.

Note: See also section 39.

38 Commissioner may require compliance reports

(1) This section applies to the following services:
  (a) a Tier 1 designated internet service;
  (b) a Tier 2 designated internet service;
  (c) an enterprise DIS;
  (d) a high impact generative AI DIS;
  (e) a machine learning model platform service.

Minimum requirements—generally

(2) The Commissioner may, by notice, require the provider of a service to prepare and give the Commissioner a report that:
  (a) specifies the steps that the provider has taken, including measures and controls the provider has implemented, to comply with applicable minimum compliance measures in this Part;
  (b) includes confirmation from the provider that the steps, measures and controls are appropriate, including reasonable supporting details and evidence; and
  (c) where applicable for the relevant designated internet service, such other details as specified in subsections (7) and (8).

(3) However, the Commissioner may not request a report under this section in respect of a designated internet service:
  (a) at any time prior to the first anniversary of the commencement of this industry standard; and
  (b) without limiting paragraph (a), more than once in any 12 month period.

(4) The notice may require the report to be in a specified form.

(5) The provider must give the report to the Commissioner within the period specified in the notice.

Note: See also section 39.

(6) A compliance report may relate to 2 or more services.

Further minimum requirements —machine learning model platform service

(7) Without limiting subsection (2), the provider of a machine learning model platform service must ensure that any report required by this section for a calendar year:
  (a) specifies:
    (i) the volume of child sexual exploitation material and pro-terror material identified by the provider in relation to the service in the calendar year, where it is technically feasible for the provider to identify such material; and

(ii) the number of models made available through the service during the calendar year for which it is reasonably foreseeable that the model could be used to generate CSEM or pro-terror material;

(b) specifies the way in which the details and materials under paragraph (a) (if any) were identified;

(c) includes details of the action taken by the provider in the calendar year in respect of the details and materials identified as mentioned in paragraph (a); and

(d) specifies the average number of Australian monthly active users of the service in the calendar year, and how that number was worked out.

Example: For paragraph (b): identification through reports made to the provider, hashing or through other measures and controls implemented by the provider.

Further minimum requirements—Tier 1 designated internet service, end-user managed hosting service, high impact generative AI DIS

(8) Without limiting subsection (2), the provider of a Tier 1 designated internet service, end-user managed hosting service or high impact generative AI DIS must ensure that the compliance report:

(a) specifies the volume of child sexual exploitation material and pro-terror material identified by the provider in relation to the service;

(b) specifies the manner in which the materials under paragraph (a) (if any) were identified;

(c) includes details of the action taken by the provider in respect of materials identified under paragraph (a);

(d) specifies the average number of Australian monthly active users of the service in the prior 12 month period, and how that number was worked out.

Example: For paragraph (b): identification through reports made to the provider, hashing or through other measures and controls implemented by the provider.

Further minimum requirements—end-user managed hosting service and high impact generative AI DIS

(9) Without limiting subsections (2) and (8), the provider of an end-user managed hosting service or a high impact generative AI DIS must ensure that the compliance report sets out:

(a) details of any limitations on the service or the provider to identify, assess or take action in respect of class 1A material and class 1B material; and

(b) where relevant, a description of the design and technology features of the service giving rise to the limitations under (a); and

(c) the impact of such limitations on the matters specified in paragraphs (8)(a), (b) and (c).

39 Extension of reporting periods

The Commissioner may, on application, extend the period within which a provider must give the Commissioner a report, certificate or notification under this Division, and may do so before or after the period has expired.

Part 5—Miscellaneous

40 Complaint resolution arrangements

(1) This section applies to a designated internet service if this industry standard requires the provider to make provision in respect of complaints by end-users in Australia of the service.

(2) If a complaint in relation to the service is made by an end-user, the provider must:
   (a) investigate the complaint; and
   (b) notify the complainant of the outcome of the investigations and the action proposed by the provider to in consequence of the investigation.

(3) Subsection (2) does not apply if:
   (a) the provider believes on reasonable grounds that the complaint was frivolous or vexatious or otherwise not made in good faith; or
   (b) the matter the subject of the complaint is being investigated, or has been investigated, by the Commissioner under Division 5 of Part 3 of the Act.

Commented [A56]: This section does not make sense for pre-assessed and other services (e.g. Tier 3) that don't have any obligations under the standards. To make it clearer to providers whether this applies to them or not, this should be included within the relevant section (re complaints) and not remain a standalone section.

41 Record-keeping

(1) The section applies to all designated internet services.

(2) The provider of a service must keep records that set out the actions that the provider has taken to comply with this industry standard.

(3) The provider must keep the records for at least 2 years after the end of the calendar year during which the action was taken.

Commented [A57]: This MCM does not make sense for pre-assessed and other services (e.g. Tier 3) that don't have any obligations under the standards. What records are they keeping if they have no compliance obligations?

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

I, Julie Inman Grant, eSafety Commissioner, determine the following industry standard.

Dated

DRAFT ONLY—NOT FOR SIGNATURE

Julie Inman Grant
eSafety Commissioner

Contents [table of contents removed]

Part 1—Preliminary

1 Name

This is the *Online Safety (Relevant Electronic Services—Class 1A and Class 1B Material) Industry Standard 2024*.

2 Commencement

This industry standard commences on the day that is 6 months after the later of:

(a) the day after the day on which it is registered under the Act; and

(b) the day after the day on which it is registered under the *Legislation Act 2003*.

3 Authority

This industry standard is determined under section 145 of the *Online Safety Act 2021*.

4 Object of this industry standard

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

Commented [A58]: Throughout both standards, this reference and similar references should always include "and/or". It is impossible and also not necessary for all providers captured to provide all of the above. This comment applies wherever these terms appear.

5 Application of this industry standard

(1) This industry standard applies to a relevant electronic service, wherever it is provided from, but only so far as it is provided to end-users in Australia.

Commented [A59]: The extraterritorial application of the OSA is already dealt with in the OSA, including through the definition of Australians. It is unnecessary and confusing to insert additional qualifiers in the standards. If absolutely necessary, include in guidance.

(2) If:

(a) this industry standard applies to a relevant electronic service; and

(b) another industry standard, or an industry code, applies to the service; and

(c) the service's predominant functionality is more closely aligned with the other industry standard or the industry code;

this industry standard does not apply to the service.

Commented [A60]: Refer to our and DIGI submissions. At the very minimum this should reference "functionality **or** purpose"

Part 2—Interpretation

Note: A number of expressions used in this industry standard are defined in the Act, including the following:
(a) child;
(b) class 1 material;
(c) class 2 material;
(d) Classification Board;
(e) Commissioner;
(f) computer game;
(g) consent;
(h) electronic service;
(i) material;
(j) parent;
(k) posted;
(l) publication;
(m) relevant electronic service;
(o) removed;
(o) service.

6 General definitions

Definitions

(1) In this industry standard:

acceptable use policy, for a relevant electronic service, means the provisions of the terms of use for the service that regulate the use of the service by end-users.

account holder, for a relevant electronic service, means the person who is the counterparty to an agreement with the provider of the service for the provision of the service.

Example: A relevant electronic service may be provided to a family, where a parent or carer has the agreement with the provider of the service. The parent or carer is the account holder. Family members (including the parent or carer who is the account holder) who use the service are end-users.

Act means the *Online Safety Act 2021*.

appropriate action: see section 12.

Australian child means a child who is in Australia.

child sexual abuse material means material that:

~~(a)~~ describes, depicts, promotes or provides instruction in child sexual abuse;

~~or~~

~~(b) is known child sexual abuse material~~.

child sexual exploitation material means material that:

(a) is or includes material that promotes, or provides instruction in, paedophile activity; or

(b) is or includes:

(i) child sexual abuse material; or

Commented [A61]: Refer to our and DIGI submissions with respect to all definitions that relate to material/Classification Scheme.

Commented [A62]: Known CSAM is defined separately.

(ii) exploitative or offensive descriptions or depictions involving a person who is, appears to be or is described as a child; or

(c) describes or depicts, in a way that is likely to cause offence to a reasonable adult, a person who is, appears to be or is described as a child (whether or not the person is engaged in sexual activity);

and, in the case of a publication, also includes material that is or includes gratuitous, exploitative or offensive descriptions or depictions of:

(d) sexualised nudity; or

(e) sexual activity involving a person who is, appears to be or is described as a child.

class 1A material means class 1 material so far as it comprises:

(a) child sexual exploitation material; or

(b) pro-terror material; or

(c) extreme crime and violence material.

Note: For the definition of ***class 1 material*** see section 106 of the Act.

class 1B material means class 1 material so far as it comprises:

(a) crime and violence material (but not extreme crime and violence material); or

(b) drug-related material.

Note: For the definition of ***class 1 material*** see section 106 of the Act.

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

Note: RC is a classification.

closed communication relevant electronic service means a relevant electronic service the primary functionality of which is to enable an end-user in Australia:

(a) to create a list of other end-users of the service (***target end-users***); and

(b) to access and communicate with target end-users on that list;

where the first end-user obtained ~~has~~ the target end-users' contact details otherwise than from the service, but does not include a service that is able to recommend target end-users to end-users in Australia based on interests or connections common to the end-users.

compliance report means a report required by section 37 or under subsection 38(2).

crime and violence material, in relation to a computer game, means material that is accessible in ~~is~~ a computer game and that, without justification:

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in, or promotion of, matters of crime or violence; or

(b) is or includes depictions of bestiality or similar practices; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality,

Commented [A63]: Refer to our submission.

Commented [A64]: Refer to our submission re the inclusion of CSP services in the definition of closed communication RES.

Commented [A65]: "material that is a computer game" does not make sense.

decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes depictions of violence that:
  (i) have a very high degree of impact; and
  (ii) are excessively frequent, prolonged, detailed or repetitive; or

(e) is or includes depictions of cruelty or realistic violence that:
  (i) have a very high degree of impact; and
  (ii) are very detailed; or

(f) is or includes depictions of actual sexual violence; or

(g) is or includes depictions of implied sexual violence related to incentives or rewards.

crime and violence material, in relation to a publication, means material that is, or is included in, the publication and that, without justification:

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in matters of crime or violence; or

(b) is or includes realistic depictions of bestiality; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes gratuitous, exploitative or offensive descriptions or depictions of violence that:
  (i) have a very high degree of impact; and
  (ii) are excessively frequent, emphasised or detailed; or

(e) is or includes gratuitous, exploitative or offensive descriptions or depictions of cruelty or real violence that:
  (i) have a very high degree of impact; and
  (ii) are very detailed; or

(f) is or includes gratuitous, exploitative or offensive descriptions or depictions of sexual violence.

crime and violence material, ~~in relation to material~~ that is not a computer game or a publication, means material that, without justification:

(a) promotes, incites or instructs in matters of crime or violence, or is or includes detailed instruction in matters of crime or violence; or

(b) is or includes depictions of bestiality or similar practices; or

(c) depicts, expresses or otherwise deals with matters of crime, cruelty or violence in such a way that it offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that it should be classified RC; or

(d) is or includes gratuitous, exploitative or offensive depictions of violence that:
  (i) have a very high degree of impact; or
  (ii) are excessively frequent, prolonged or detailed; or

(e) is or includes gratuitous, exploitative or offensive depictions of cruelty or real violence that:
  (i) have a very high degree of impact; and
  (ii) are very detailed; or
(f) is or includes gratuitous, exploitative or offensive depictions of sexual violence.

CSP means a carriage service provider as defined in section 87 of the *Telecommunications Act 1997*.

dating service means a relevant electronic service the primary functionality of which is:
(a) to solicit, offer, promote or provide access to dating, relationship, compatibility, matrimonial, social or romantic referral services; and
(b) to enable end-users to communicate with other end-users online;
but does not include such a service to the extent that its functionality is to connect end-users who offer their services for payment.

Note: Examples of services for payment are escort or ~~prostitute~~ ~~services~~sex work.

Commented [A66]: As far as we understand, the sex workers industry rejects this language.

development program means a program required by section 23.

drug means a chemical, compound, or other substance or thing, that is included in Schedule 4 of the *Customs (Prohibited Imports) Regulations 1956*.

drug-related material, in relation to a computer game, means material that, without justification:
(a) depicts the unlawful use of drugs in connection with incentives or rewards; or
(b) depicts interactive, detailed and realistic use of drugs, being unlawful use; or
(c) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or
(b) is or includes detailed instruction in the unlawful use of drugs.

drug-related material, in relation to a publication, means material that, without justification:
(a) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or
(b) is or includes detailed instruction in the unlawful use of drugs.

drug-related material, ~~in relation to material~~ that is not a computer game or a publication, means material that, without justification:
(a) depicts, expresses or otherwise deals with matters of drug misuse or addiction in such a way that the material offends against the standards of

morality, decency and propriety generally accepted by reasonable adults to the extent that the material should be classified RC; or

(b) is or includes detailed instruction in the unlawful use of drugs; or

(c) is or includes material promoting the unlawful use of drugs.

end-user, of a relevant online service, means a natural person who uses the service.

Example: A relevant electronic service may be provided to a family, where a parent or carer has the agreement with the provider of the service. The parent or carer is the account holder. Family members (including the parent or carer who is the account holder) who use the service are end-users.

enforcement authority means:

(a) a police force or other law enforcement authority; or

(b) an organisation (including a non-government organisation) the functions of which include receiving reports of child sexual exploitation material or pro-terror material and facilitating making those reports to law enforcement authorities.

enterprise relevant electronic service means a relevant electronic service:

(a) the account holder for which is an organisation ~~(and not an individual)~~; and

(b) the primary purpose ~~functionality~~ of which is to enable the account holder, in accordance with the terms of use for the service, to make the service available to a specified class of persons to facilitate communications between those persons; and

(b) that is of a kind that is usually acquired by account holders for the purpose mentioned in paragraph (b).

exploitative, in relation to a description or depiction of an event, means that the description or depiction:

(a) appears intended to debase or abuse, for the enjoyment of readers or viewers, the person or entity described or depicted; and

(b) has no moral, artistic or other value.

extreme crime and violence material, in relation to a computer game, means crime and violence material ~~that is crime and violence material in relation to a computer game~~ where, without justification, the impact of the material is extreme because:

(a) the material is more detailed; or

(b) the material is realistic rather than stylised; or

(c) the game is highly interactive; or

(d) the gameplay links incentives or rewards to high impact elements of the game; or

(e) for any other reason.

extreme crime and violence material, in relation to a publication, means crime and violence material ~~that is crime and violence material in relation to a publication~~ where, without justification, the impact of the material is extreme

Commented [A67]: Refer to our submission.

Commented [A68]: Refer to our submission.

Commented [A69]: It is unclear why functionality is relevant here. The purpose of offering the service is the key question. The functionality will be the same whether it is being offered to a natural person, or an organisation.

because of the emphasis, tone, frequency, context and detail of the relevant elements of the publication and other factors that heighten impact.

extreme crime and violence material, ~~in relation to material~~ that is not a computer game or a publication, means crime and violence material where, without justification, the impact of the material is extreme because:

(a) the material is more detailed; or

(b) the material is highly interactive; or

(c) the relevant depictions in the material are realistic, prolonged or repeated; or

(d) for any other reason.

gaming service with communications functionality means a relevant electronic service the primary functionality of which is:

(a) to enable end-users in Australia to play online games with other end-users; and

(b) to enable sharing of user-generated URLs, hyper-linked text, images or videos between end-users;

but does not include

(c) a closed communication relevant electronic service; or

(d) a gaming service with limited communications functionality; or

(e) a service that limits the sharing of user-generated material between end-users to any or all of the following:

(i) in-game images or footage;

(ii) user-generated designs (such as environments and artwork);

(iii) virtual objects or maps;

(iv) pre-selected messages;

(v) non-hyper-linked text that is subject to automated filtering technology; or

(vi) ephemeral voice interactions.

gaming service with limited communications functionality means a relevant electronic service, other than a closed communication relevant electronic service, the primary functionality of which is to enable end-users in Australia to play online games with other end-users, without enabling the sharing of user-generated URLs, hyper-linked text, images or videos between end-users (other than material of a kind referred to in paragraph (e) of the definition of gaming service with communications functionality in this subsection).

industry code has the meaning given in section 132 of the Act.

justification: see subsection (2).

known child sexual abuse material means material that:

(a) is or includes images (either still images or video images); and

(b) has been verified as child sexual abuse material by a governmental (including multi-lateral) or non-governmental organisation;

Commented [A70]: Refer to our and DIGI submissions.

(i) the functions of which are or include combating child sexual abuse or child sexual exploitation; and

(ii) in the case of a non-governmental organisation—that is generally recognised as expert or authoritative in that context; and

(c) is recorded on a database that:

(i) is managed by an organisation of a kind described in paragraph (b); and

(ii) is made available to government agencies, enforcement authorities and providers of relevant electronic services for the purpose of ~~their~~ using technological means to detect or manage child sexual abuse material on relevant electronic services.

Example: An example of a database referred to in paragraph (c) is the database managed by the National Center for Missing & Exploited Children.

known pro-terror material means material that has been verified as pro-terror material.

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

That List was accessible, on the registration of this industry standard, at https://www.un.org/securitycouncil/content/un-sc-consolidated-list.

Note 2: Material may, for example, be verified as a result of a decision of the Classification Board. Material may also be verified by using tools provided by independent organisations that are recognised as having expertise in counter-terrorism. Examples of these organisations include Tech against Terrorism and the Global Internet Forum to Counter Terrorism.

offensive: see subsection (3).

open communication relevant electronic service means a relevant electronic service the primary functionality of which is:

(a) to enable end-users in Australia to view, search for or communicate with other end-users (***target end-users***) on the service without knowing the target end-users' contact details; or

(b) to recommend target end-users to end-users in Australia, based on interests or connections common to the end-users.

To avoid doubt, it includes a relevant electronic service that enables an end-user to invite, through use of an internet link, another end-user to communicate with the first end-user.

pre-assessed relevant electronic service means each of the following:

(a) a closed communication relevant electronic service;

(b) a dating service;

(c) a gaming service with communications functionality;

(d) an open communication relevant electronic service.

pro-terror material means:

(a) material that:

Commented [A71]: Refer to our and DIGI submissions. Also, this definition moves the absolute standard of inclusion on the UNSC list (in RES Code) into a note and as a 'may'. This produces uncertainty re 'verification'. For example, in the current Gaza conflict, there may be material that is 'verified' by Arabic states as pro-terror that would not meet that definition. Similarly, not all material 'verified' by Israel may meet that definition.

Commented [A72]: Refer to our and DIGI submissions.

(i) directly or indirectly counsels, promotes, encourages or urges the doing of a terrorist act; or

(ii) directly or indirectly provides instruction in the doing of a terrorist act; or

(iii) directly praises the doing of a terrorist act in circumstances where there is a substantial risk that the praise might have the effect of a leading a person (regardless of the person's age or any mental impairment that the person might suffer) to engage in a terrorist act; or

(b) material that is known pro-terror material.

However, material accessible using a relevant electronic service is not pro-terror material if its availability on the service can reasonably be taken to be part of public discussion, public debate, entertainment or satire.

provide a relevant electronic service includes make the service available.

provider, in relation to a ***CSP*** ~~***telephony***~~ relevant electronic service, has the meaning given to ***carriage service provider*** in section 87 of the *Telecommunications Act 1997*.

RC means the "Refused Classification" classification under the National Classification Code.

risk assessment means an assessment of a kind required by subsection 8(1).

risk profile, for a relevant electronic service, means the risk profile of the service worked out under subsection 8(7).

sexual activity is not limited to sexual intercourse.

store: material is ***stored on a relevant electronic service*** if it is:

(a) in storage used for the service; or

(b) accessible through or using the service.

CSP ~~***telephony***~~ ***relevant electronic service*** means a relevant electronic service provided by a carriage service provider as defined in ~~short messaging service (SMS) or a multimedia messaging service (MMS) provided over a public mobile telecommunications service as defined in in sub~~section 87~~32(1)~~ of the *Telecommunications Act 1997*.

Commented [A73]: Refer to our submission.

terrorist act has the meaning given by section 100.1(1) of the *Criminal Code* (no matter where the action occurs, the threat of action is made or the action, if carried out, would occur).

terms of use, for a relevant electronic service, means the provisions of the agreement under which the service is provided and includes anything that may reasonably be regarded as the equivalent of terms of use.

Commented [A74]: Refer to our submission.

Tier 1 relevant electronic service means a relevant electronic service:

(a) that is a Tier 1 relevant electronic service under paragraph 8(7)(a); or

(b) that is determined under subsection 8(9) to be a Tier 1 relevant electronic service.

Tier 2 relevant electronic service means a relevant electronic service that is a Tier 2 relevant electronic service under paragraph 8(7)(b).

violence means an act of violence or an obvious threat of an act of violence.

young Australian child means Australian child who is under 16.

Justification

(2) For this industry standard, in determining whether material is without justification, the matters to be taken into account include:

(a) the standards of morality, decency and propriety generally accepted by reasonable adults; and

(b) the literary, artistic or educational merit (if any) of the material; and

(c) the general character of the material, including whether it is of a medical, legal or scientific character; and

(d) the persons or class of persons to or amongst whom it is published or is intended or likely to be published.

Offensive material

(3) The question whether material is offensive for the purposes of this industry standard is to be determined in accordance with the Act, including section 8 of the Act.

7 Technical feasibility

In considering whether it is or is not technically feasible for the provider of a relevant electronic service to take a particular action, the matters to be taken into account include:

(a) the expected financial cost to the provider of taking the action; and

(b) whether it is reasonable to expect the provider to incur that cost, having regard to the level of the risk to the online safety of end-users in Australia of not taking the action.

Commented [A75]: Definition of Tier 3 is missing but used later-on.

Commented [A76]: Refer to our and DIGI submissions in relation to context and Classification Scheme.

Commented [A77]: Refer to our and DIGI submissions.

Part 3—Risk assessments and risk profiles

8 Requirement to carry out risk assessments and determine risk profiles of relevant electronic services

Risk assessments to be carried out

(1) The provider of a relevant electronic service must, at the times required by and in accordance with this Part, carry out an assessment of the risk that class 1A material or class 1B material:

(a) will be generated or accessed by, or distributed by or to, end-users in Australia using the service; and

(b) will be stored on the service.

Note: See also paragraph 33(b).

Timing of risk assessments

(2) If the provider of the service was providing the service before the commencement of this industry standard, the risk assessment must be carried out as soon as practicable after, but no later than 6 months after, the commencement of this industry standard.

(3) Subsection (2) does not apply if a risk assessment that met the requirements of this Part prior to the service being made available in Australia ~~had been carried out in respect of the service within 6 months before the commencement of this industry standard~~.

(4) A person must not start to provide a relevant electronic service to an end-user in Australia unless a risk assessment of the service has been carried out in accordance with this Part within 6 months before the person started to provide the service.

(5) The provider of a relevant electronic service must not make a material change to the service unless:

(a) a risk assessment of the service, as proposed to be changed, has been carried out in accordance with this Part; or

(b) the change will not materially increase the risk of class 1A material or class 1B material being accessed by, or distributed to, end-users in Australia using the service, or being stored on the service.

Certain services exempt from risk assessment requirements

(6) Subsections (1) and (4) do not apply to any of the following:

(a) an enterprise relevant electronic service;

(b) a gaming service with limited communications functionality;

(c) a pre-assessed relevant electronic service;

(d) a relevant electronic service that is determined under subsection (9) to be a Tier 1 relevant electronic service.

Commented [A78]: Refer to our submission re risk mitigation.

Commented [A79]: See our submission.

Commented [A80]: Refer to DIGI submission for feedback on inert material.

Commented [A81]: Refer to our submission.

Commented [A82]: Refer to DIDI submission for a discussion of the test for changes.
Also, adding 'material' aligns with the reporting obligations further below.

Commented [A83]: Refer to our submission.

Note: However, subsection (1) applies to the provider of a relevant electronic service mentioned in this subsection if it makes a material change to the service and this change would increase the risk of class 1A or class 1B material being accessed or generated by, or distributed to end-users in Australia using the service ~~the service is materially changed~~.

Risk profiles of relevant electronic services

(7) The risk profile of a relevant electronic service is worked out as follows:

(a) if the risk that class 1A material or class 1B material will be solicited or accessed by, or distributed to, end-users in Australia using the service, or will be stored on the service, is high—the service is a Tier 1 service;

(b) if the risk that class 1A material or class 1B material will be solicited or accessed by, or distributed to, end-users in Australia using the service, or will be stored on the service, is medium—the service is a Tier 2 service.

(c) if the risk that class 1A material or class 1B material will be solicited or accessed by, or distributed to, end-users in Australia using the service, or will be stored on the service, is low—the service is a Tier 3 service.

Commented [A84]: See comment at definitions - Tier 3 not defined.

(8) The provider of a relevant electronic service that conducts a risk assessment of the service must, on completion of the assessment, determine, in accordance with subsection (7), what the risk profile of the service is.

(9) However, the provider of a relevant electronic service may, at any time, without having conducted a risk assessment, determine that the risk profile of the service is Tier 1.

Note: See also paragraph 33(1)(b).

9 Methodology, risk factors and indicators to be used for risk assessments and risk profile determinations

Requirement for plan and methodology

(1) If the provider is required by this Part to carry out a risk assessment for a service, the provider must formulate in writing a plan, and a methodology, for carrying out the assessment that ensure that the risks mentioned in subsection 8(1) in relation to the service are accurately assessed.

(2) The provider must ensure that the risk assessment is carried out in accordance with the plan and methodology.

(3) The provider must ensure that a risk assessment is carried out by persons with the relevant skills, experience and expertise.

Forward-looking analyses of likely changes

(4) As part of a risk assessment carried out as required by this Part, the provider must undertake a forward-looking analysis of:

(a) likely changes to the internal and external environment in which the service operates or will operate, including likely changes in the functionality of, or the scale of, the service; and

(b) the impact of those changes on the ability of the service to meet the object of this industry standard.

Note: For the object of this industry standard see section 4.

Matters to be taken into account

(5) Without limiting subsection (1), the methodology for the conduct of a risk assessment must specify the principal matters to be taken into account in assessing relevant risks, which must include the following, so far as they are relevant to the service:

(a) the predominant functionality of the service;

(b) the extent to which material posted on or distributed using the service will be available to end-users of the service in Australia;

(c) the manner in which material is created or contributed to in connection with the service;

(d~~c~~) the terms of use for the service;

(e~~d~~) the terms of arrangements under which the provider acquires content to be made available on the service;

(f~~e~~) the ages of end-users and likely end-users of the service;

(g~~f~~) the outcomes of the analysis conducted as required by subsection (4);

(h~~g~~) safety by design principles incorporated or relied upon during the design or operation of the service, including guidance and tools published or made available by a government agency or a foreign or international body;

(i~~h~~) the risk to the online safety of end-users in Australia in relation to synthetic material generated by artificial intelligence.

Note 1: Arrangements referred to in paragraph (d) may include provisions that, if complied with, will reduce the risk that class 1A material and class 1B material will be made available through the service.

Note 2: Safety by design guidance or principles in paragraph (h) may be derived by government, international organisations or technology service providers. Examples of agencies mentioned in paragraph (g) are the Commissioner or the Digital Trust & Safety Partnership Safe Framework.

Commented [A85]: Suggest also including DIS s9(5)(b): "the manner in which material is created or contributed to in connection with the service". This could provide flexibility, and capture context-specific communications.

Commented [A86]: Refer to our submission.

Commented [A87]: Refer to our submission.

10 Documenting risk assessments and risk profiles

(1) As soon as practicable after determining the risk profile of a relevant electronic service, the provider of the service must record in writing:

(a) details of the determination; and

(b) details of the conduct of any related risk assessment;

sufficient to demonstrate that the determination and the risk assessment were made or carried out in accordance with this Part.

(2) The record must include the reasons for the results of the assessment and the determination of the risk profile.

Note: See also paragraph 33(b).

Part 4—Online safety compliance measures

Division 1—Preliminary

11 This Part not exhaustive

This Part does not prevent the provider of a relevant electronic service from taking measures, in addition to and not inconsistent with those required by this Part, to improve and promote online safety for Australians.

12 What is appropriate action?

Commented [A88]: Refer to DIGI submission.

In determining whether action taken or proposed in relation to a relevant electronic service as required by this industry standard is appropriate, the matters to be taken into account include:

(a) the extent to which the action achieves the object of this industry standard in relation to the service; and

(b) if the action relates to a breach of applicable terms of use of a relevant electronic service, or community standards, in relation to class 1A material or class 1B material:

(i) the nature of the material and the extent to which the breach is inconsistent with online safety for end-users in Australia; and

(ii) the extent to which the action will or may reasonably be expected to reduce or manage the risk that the service will be used to solicit, access, communicate or store class 1A material or class 1B material; and

(iii) whether the proposed action is proportionate to the level of risk to online safety for end-users in Australia from the material being accessible through the service.

Note: For the object of this industry standard see section 4.

13 Index of requirements for relevant electronic services

The following table sets out the provisions of this Part applicable to providers of relevant electronic services.

Item	For this kind of relevant electronic service …	the applicable provisions of this Part are…
1	all relevant electronic services	sections 33 and 34
2	pre-assessed relevant electronic services	(a) the provisions listed in item 1 (b) sections 14 to 28, 30, 31 and 37
3	closed communication relevant electronic services	(a) the provisions listed in items 1 and 2 (b) section 35
4	dating services	(a) the provisions listed in items 1 and 2 (b) sections 32, 35 and 38
5	enterprise relevant electronic services	(a) the provisions listed in item 1

Commented [A89]: It is unclear how these three compliance obligations will apply to pre-assessed services pre-assessed services given they are not required to complete a risk assessment to comply with the standard. Some RES will have no obligations other than where they make a material change to the service, and that is not acknowledged in the Standard.

Item	For this kind of relevant electronic service …	the applicable provisions of this Part are…
		(b) section 14 and subsection 38(1)
6	gaming services with communications functionality	(a) the provisions listed in items 1 and 2 (b) subsection 38(2)
7	gaming services with limited communications functionality	the provisions listed in item 1
8	open communication relevant electronic services	(a) the provisions listed in items 1 and 2 (b) sections 32 and 35
9	~~telephony~~ CSP relevant electronic service	(a) the provisions listed in item 1 (b) sections 14 to 17, 24 to 28, 30, 31 and section 3~~7~~8
10	Tier 1 relevant electronic service	(a) the provisions listed in item 1 (b) sections 14 to 28, 30 and 31
11	Tier 2 relevant electronic service	(a) the provisions listed in item 1 (b) sections 14 to 19, 24 to 28, 30, 31 and subsection 38(2)
12	Tier 3 relevant electronic service	the provisions listed in item 1

Commented [A90]: Not according to section 37. Section 38 lists these providers.

Division 2—Compliance measures

14 Terms of use

(1) This section applies to the following:
- (a) an enterprise relevant electronic service;
- (b) a pre-assessed relevant electronic service;
- (c) a CSP~~telephony~~ relevant electronic service;
- (d) a Tier 1 relevant electronic service;
- (e) a Tier 2 relevant electronic service.

Provisions to be included in terms of use

(2) The provider of a service must include in the terms of use for the service provisions:
- (a) imposing an obligation on the account holder of the service to ensure that the service is not used, whether by the account holder, or by an end-user in Australia, to solicit, access, distribute or store class 1A material or class 1B material; and
- (b) giving rights for the provider to do any of the following if the service is used to solicit, access, distribute or store class 1A material or class 1B material:
  - (i) suspend the provision of the service to a specified end-user of the service for a specified period;
  - (ii) impose specified restrictions on the use of the service by a specified end-user of the service for a specified period;
  - (iii) terminate the agreement for the provision of the service.

Enforcement of terms of use

(3) If the provider of a relevant electronic service becomes aware of a breach of the obligation mentioned in paragraph (2)(a), the provider must enforce its contractual rights in respect of the breach in an appropriate way that is proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach.

(4) In proceedings in respect of a contravention of subsection (3), the provider bears the evidential burden of establishing:
- (a) the action it took to enforce the rights; and
- (b) that the action that it took was appropriate and proportionate, as referred to in subsection (3).

Note: For appropriate action see also section 12.

15 Notification of child sexual exploitation material and pro-terror material

(1) This section applies to the following:
- (a) a gaming service with limited communication functionality;

Commented [A91]: Refer to our submission.

Commented [A92]: Refer to our submission.

Commented [A93]: Refer to our submission.

Commented [A94]: Refer to our submission.

Commented [A95]: Refer to DIGI submission for a discussion of available databases, authorities, context etc. required for the detection of such material and notification to authorities.

(b) a pre-assessed relevant electronic service;
(c) a CSP ~~telephony~~ relevant electronic service;
(d) a Tier 1 relevant electronic service; and
(e) a Tier 2 relevant electronic service.

(2) If the provider of a service:
(a) identifies child sexual exploitation material, or pro-terror material, on the service; and
(b) believes in good faith that the material affords evidence of a serious and immediate threat to the life or physical safety of a person in Australia;
the provider must, as soon as practicable, report the matter to an enforcement authority, or otherwise as required by law.

(3) If the provider of a service:
(a) identifies child sexual abuse ~~exploitation~~ material on the service; ~~and~~
(b) believes in good faith that the material is reasonably likely to be ~~not known~~ child sexual abuse ~~exploitation~~ material; and
(c) believes in good faith that the material is not known child sexual abuse material;
the provider must, as soon as practicable, notify an organisation of a kind referred to in paragraph (b) of the definition of known child sexual abuse ~~exploitation~~ material in subsection 6(1).

Commented [A96]: Child sexual exploitation material (being the broadest category) is already dealt with in (2) so this section should be limited to CSAM. Depictions of child-like characters (e.g. erotic manga) should not be reported to NCMEC but under the proposed eSafety drafting, this would be a requirement. This requirement should really only capture actual abuse material.

(4) If the provider of a service:
(a) identifies pro-terror material on the service; and
(b) believes in good faith that the material is not known pro-terror material;
the provider must, as soon as practicable, notify an organisation that verifies material as pro-terror material.

Note: See the definition of ***pro-terror material*** in subsection 6(1).

(5) Subsections (2), (3) and (4) are in addition to any other applicable law.

16 Systems and processes for responding to breaches of terms of use or community standards: class 1A material

Commented [A97]: Refer to our submission.

(1) This section applies to the following:
(a) a pre-assessed relevant electronic service;
(b) a CSP ~~telephony~~ relevant electronic service;
(c) a Tier 1 relevant electronic service;
(d) a Tier 2 relevant electronic service.

(2) The provider of a service must implement systems and processes that ensure that, if the provider becomes aware that:
(a) there is or has been a breach of an obligation under the terms of use for the service in respect of class 1A material, including a breach of an obligation to comply with acceptable use policies; or

(b) there is or has been a breach, in Australia, involving the service, of community standards in respect of class 1A material;

the provider takes appropriate action to ensure that:

(c) the breach, if it is continuing, ceases; and

(d) the risk of further such breaches is minimised.

(3) Without limiting subsection (2), the systems, processes must include ones under which the provider:

(a) reviews reports by end-users of the service in Australia that class 1A materials are accessible using the service; and

(b) appropriately prioritises those reports and, if necessary, escalates them to senior management personnel of the provider for action.

Commented [A98]: Refer to our submission.

17 Responding to breaches of terms of use or community standards: class 1A material

Commented [A99]: The corresponding section in the DIS Standard (s17) only addresses "CSEM and pro-terror material" . It is not apparent why RES ought to take a different approach. We recommend aligning the RES standard with the DIS standard in this respect.

Note: For breaches in respect of class 1B material see section 25.

(1) This section applies to the following:

(a) a pre-assessed relevant electronic service;

(b) a CSP~~telephony~~ relevant electronic service;

(d) a Tier 1 relevant electronic service;

(d) a Tier 2 relevant electronic service.

(2) If the provider of a service becomes aware that:

(a) there is or has been a breach of an obligation under the terms of use for the service in respect of class 1A material, including a breach of an obligation to comply with acceptable use policies; or

(b) there is or has been a breach, in Australia, involving the service, of community standards in respect of class 1A material;

the provider must:

(c) as soon as practicable, remove the material, or cause the material to be removed, from the service unless it is not technically feasible for the provider to do so; and

(d) take appropriate action to ensure that:

(i) the service no longer permits access to or distribution of the material; and

(ii) the breach, if it is continuing, ceases; and

(iii) the risk of further such breaches is minimised.

Note: For appropriate action see section 12.

(3) Without limiting what is appropriate action, appropriate action may include exercising, in a way that is proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach, any of the provider's contractual rights under the terms of use for the service in relation to the breach.

Note: For contractual rights required to be included in terms of use see paragraph 14(1)(b).

(4) If the provider of a service becomes aware that an end-user in Australia of the service has breached obligations or standards mentioned in paragraph (2)(a) in respect of child sexual exploitation material or pro-terror material, the provider must ensure that all the child sexual exploitation material or pro-terror material is removed from the service as soon as practicable after the provider becomes aware the breach.

Commented [A100]: Refer to our submission.

(5) Subsection (4) does not affect paragraph (2)(c) and does not apply if it is not technically feasible for the provider to remove the material from the service.

18 Resourcing trust and safety functions

(1) This section applies to the following:
- (a) a pre-assessed relevant electronic service;
- (b) a Tier 1 relevant electronic service;
- (c) a Tier 2 relevant electronic service.

(2) The provider of a relevant electronic service must have and implement, in respect of the service, management, supervision and internal reporting arrangements to ensure that at all times the provider:
- (a) complies with the requirements of this industry standard; and
- (b) can otherwise effectively supervise the online safety of the service.

Note These arrangements may include duties and responsibilities for personnel, and systems, processes and technologies.

(3) The provider of a relevant electronic service must have, or have access to, sufficient personnel who have the skills, experience and qualifications needed to ensure that the provider complies with the requirements of this industry standard at all times.

19 Safety features and settings

(1) This section applies to the following:
- (a) a pre-assessed relevant electronic service;
- (b) a Tier 1 relevant electronic service; and
- (c) a Tier 2 relevant electronic service.

(2) Before the provider of the service makes a material change to the service, the provider must:
- (a) carry out an assessment of the kinds of features and settings that could be incorporated into the service to minimise the risk that class 1A material:
  - (i) will be accessed by, or distributed to, end-users in Australia using the service; or
  - (ii) will be stored on the service; and
- (b) determine, on the basis of the assessment, the most appropriate and effective features and settings for the service; and
- (c) ensure that the service as so changed incorporates at all times the features and settings so determined.

(3) Subsections (4), (5), (6) and (7) do not limit subsection (2) and apply whether or not a material change is made or proposed to the service.

Open communication relevant electronic services and Tier 1 relevant electronic services

(4) In the case of:

(a) an open communication relevant electronic service; or

(b) a Tier 1 relevant electronic service;

the provider must ensure that:

(c) if the service allows the sending of messages between end-users—it has tools and settings that allow end-users in Australia to block messages from other end-users; and

(d) if the service displays, or allows for the display of, an end-user's online status—it has tools and settings that an end-user in Australia can use to prevent the display or communication of the end-user's online status; and

(e) if the provider allows young Australian children to become account-holders or end-users of the service—it has tools and settings that prevent end-users who are over 18 from using the service to contact a young Australian child unless with the consent of the child's parent or guardian;

(f) the account of a young Australian child with the service is private by default; and

(g) the location of a young Australian child who is an end-user of the service is not available to end-users of the service unless with the consent of the child's parent or guardian.

Commented [A101]: what does that mean - geo location, city, no location at all?

Dating services

(5) The provider of a dating service must ensure that the tools and settings for the service:

(a) allow an end-user of the service to block messages from another end-user of the service; and

(b) do not permit a person to become an end-user of the service unless the person is registered with the service as an end-user; and

(c) do not permit a person to register with the service as an end-user unless the person provides the person's phone number, email address or other identifier.

(6) If a child in Australia becomes an end-user of a dating service, the provider of the service contravenes this subsection unless the provider had taken reasonable steps to ensure that children do not become end-users of the service.

Closed communication relevant electronic services

(7) The provider of a closed communication relevant electronic service must ensure that the settings for the service:

(a) do not permit a person to become an end-user of the service unless the person is registered with the service as an end-user; and

(b) do not permit a person to register as an end-user of the service unless the person provides the person's phone number, email address or other identifier.

Data retention

(8) The provider of a service must retain information provided as required by paragraph (5)(c) or (7)(b) for at least 2 years.

General information about tools and settings

(9) The provider of a service must provide information that explains the tools and settings provided as required by this section. The information:
(a) must be "in service", that is, not on a separate webpage to the webpage for the service; and
(b) must be easily accessible and easy to use; and
(c) must include or be accompanied by clear instructions on how to use the tools and settings.

Commented [A102]: Can this please be phrased in a more neutral way (retention of information provided) so that this still works when a digital ID plus associated data is used through the digital ID framework, i.e. the data will not be 'provided' anymore by the customer but only authorised to be accessed/used.

Commented [A103]: Refer to our submission.

Commented [A104]: Webpage vs website? Check for consistent terminology.

20 Detecting and removing known child sexual abuse material

(1) This section applies to the following:
(a) a pre-assessed relevant electronic service;
(b) a Tier 1 relevant electronic service.

Commented [A105]: We comment in the following on the basis of the assumption that SMS/MMS are to be excluded from the definition of closed communications RES and, therefore, pre-assessed RES (and that they are not Tier 1).

(2) The provider of a service must implement systems, processes and technologies that detect and identify known child sexual abuse material that:
(a) is stored on the service; or
(b) is accessible by an end-user in Australia using the service; or
(c) is being or has been distributed in Australia using the service.

Note: The systems, processes and technologies that the provider may use include hashing technologies, machine learning and artificial intelligence systems that scan for known child sexual abuse material.

(3) Subsection (2) does not require a provider to use a system, process or technology if it is not technically feasible for the provider to do so.

(4) The provider of a service must implement systems, processes and technologies that remove known child sexual abuse material from the service as soon as practicable after it is detected and identified.

(5) Subsection (4) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(6) If it is not technically feasible for the provider to implement a particular system, process or technology for the purposes of:
(a) detecting and identifying known child sexual abuse material as required by subsection (2); or
(b) removing known child sexual abuse material as required by subsection (4);
the provider must take appropriate alternative action.

Note: For appropriate action see section 12.

(7) This section does not affect the operation of section 22.

Note 1: For technical feasibility, see section 7.

Note 2: This section does not prevent a provider from complying with legal obligations to preserve evidence of offences.

21 Detecting and removing known pro-terror material

Commented [A106]: Refer to our and DIGI submissions for feedback in relation to pro-terror material.

(1) This section applies to the following:
(a) a pre-assessed relevant electronic service;
(b) a Tier 1 relevant electronic service.

(2) The provider of a service must implement systems, processes and technologies that detect and identify known pro-terror material that:
(a) is stored on the service; or
(b) is accessible by an end-user in Australia using the service; or
(c) is being or has been distributed in Australia using the service.

Note: The systems, processes and technologies that the provider may use include hashing technologies, machine learning and artificial intelligence systems that scan for known pro-terror material.

(3) Subsection (2) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(4) The provider of a service must implement systems, processes and technologies that remove known pro-terror material from the service as soon as practicable after it is detected and identified.

(5) Subsection (4) does not require a provider to implement a system, process or technology if it is not technically feasible for the provider to do so.

(6) If it is not technically feasible for the provider to implement a particular system, process or technology for the purposes of:
(a) detecting and identifying known pro-terror material as required by subsection (2); or
(b) removing known pro-terror material as required by subsection (4);
the provider must take appropriate alternative action.

Note: For appropriate action see section 12.

(7) This section does not affect the operation of section 22.

Note 1: For technical feasibility, see section 7.

Note 2: This section does not prevent a provider from complying with legal obligations to preserve evidence of offences.

22 Disrupting and deterring child sexual abuse material and pro-terror material

(1) This section applies to the following:
(a) a pre-assessed relevant electronic service; and
(b) a Tier 1 relevant electronic service.

(2) The provider of a service must implement systems, processes and technologies that:

(a) effectively deter end-users of the service from using the service: and

(b) effectively disrupt attempts by end-users of the service to use the service;

to create, offer, solicit, access, distribute, or otherwise make available, or store child sexual abuse material or pro-terror material (including known child sexual abuse material and known pro-terror material).

(3) Without limiting subsection (2), the systems, processes and technologies may include:

(a) hashing technologies, machine learning and artificial intelligence systems that scan for known child sexual abuse material or known pro-terror material; and

(b) systems, processes and technologies that are designed to detect key words, behavioural signals and patterns associated with child sexual abuse material.

23 Development programs

(1) This section applies to the following:

(a) a pre-assessed relevant electronic service; and

(b) a Tier 1 relevant electronic service.

for a calendar year if the average monthly number of active end-users of the service, in Australia, over the immediate previous financial year was 1,000,000 or more.

Commented [A107]: Refer to our submission.

(2) The provider of the service must establish and implement, for the calendar year, a program of investment and development activities (***development program***) in respect of systems, processes and technologies.

Note: See also section 36.

(3) A development program must include:

(a) investments and activities designed to develop systems, processes and technologies that enhance the ability of the provider, or of other providers of relevant electronic services:

(i) to detect and identify child sexual abuse material or pro-terror material (including known child sexual abuse material and known pro-terror material) on the service; and

(ii) to deter end-users of the service from using the service, and to disrupt attempts by end-users of the service to use the service, to solicit, generate, create, access, distribute or store child sexual abuse material or pro-terror material (including known child sexual abuse material and known pro-terror material); and

(iii) to reduce the risk to the online safety of end-users in Australia using the provider's service in relation to synthetic material generated by artificial intelligence; and

Commented [A108]: Otherwise the relevant providers have an obligation to make investments etc. to reduce gen AI risks in relation to online safety that are unrelated to their services.

(b) arrangements for cooperating and collaborating with other organisations in activities of the kind referred to in paragraph (a) and to enhance online safety for Australians.

Commented [A109]: Can this be discharged by participation in the industry associations' annual fora (Codes) or what exactly is the expectation?

(4) A development program may include arrangements for the provider to make available to other providers of relevant electronic services, or organisations engaged in promoting online safety for Australians, systems, processes and technologies of a kind referred to in paragraph (3)(a) (including making them available without charge).

(5) Examples of activities that may be part of a provider's development program include:

(a) joining industry organisations intended to address serious online harms; and

(b) working with the Commissioner to share information, intelligence, best practices and other information relevant to addressing categories of class 1A material or class 1B material that are relevant to the service; and

(c) collaborating with non-government or other organisations that facilitate the sharing of information, intelligence, best practices and other information relevant to addressing categories of class 1A or class 1B material that are relevant to the service.

(5) Examples of investments that may be part of a provider's development program include:

(a) procuring online safety systems and technologies for use in connection with the service, or enhancing online safety systems and technologies used in connection with the service; and

(b) conducting research into and development of online safety systems and technologies; and

(c) providing support, either financial or in kind, to organisations the functions of which are or include working to combat child sexual abuse, child sexual exploitation or terrorism.

Note: For paragraph (c), organisations can include universities, the CSIRO, the WePROTECT Global Alliance and the Global Internet Forum to Counter Terrorism (GIFCT).

24 Systems, processes and technologies for responding to breaches of terms of use or community standards: class 1B material

(1) This section applies to the following:

(a) a pre-assessed relevant electronic service;

(b) a CSP~~telephony~~ relevant electronic service;

(c) a Tier 1 relevant electronic service;

(d) a Tier 2 relevant electronic service.

(2) The provider of a service must implement systems, processes and technologies that ensure that, if the provider becomes aware that:

(a) there is or has been a breach, in Australia, of an obligation under for the terms of use for the service in respect of class 1B material, including a breach of an obligation to comply with acceptable use policies; or

(b) there is or has been a breach, in Australia, involving the service, of community standards in respect of class 1B material;

the provider takes appropriate action to ensure that:

(c) the breach, if it is continuing, ceases; and

(b) the risk of further such breaches is minimised.

(3) Without limiting subsection (2), the systems, processes and technologies must include ones under which the provider:

(a) reviews reports by end-users of the service in Australia that class 1B materials are accessible using the service; and

(b) appropriately prioritises those reports and, if necessary, escalates them to senior management personnel of the provider for action.

They must include operational guidance to provider personnel, including actions to be taken and time limits to be observed, in performing the provider's duties under this section.

25 Responding to breaches of terms of use or community standards: class 1B material

Note: For breaches in respect of class 1A material see section 17.

(1) This section applies to the following:

(a) a pre-assessed relevant electronic service;

(b) a CSP~~telephony~~ relevant electronic service;

(c) a Tier 1 relevant electronic service; and

(d) a Tier 2 relevant electronic service.

(2) If the provider of a service becomes aware that:

(a) there is or has been a breach of an obligation under the terms of use for the service in respect of class 1B material, including a breach of an obligation to comply with acceptable use policies; or

(b) there is or has been a breach involving the service, of community standards in respect of class 1B material;

the provider must, unless it is not technically feasible:

(c) as soon as practicable, remove the material, or cause the material to be removed, from the service ~~unless it is not technically feasible for the provider to do so~~; and

(d) take appropriate action to ensure that:

(i) the service no longer permits access to or distribution of the material; and

(ii) the breach, if it is continuing, ceases; and

(iii) the risk of further such breaches is minimised.

Note: For technical feasibility see section 7. For appropriate action see section 12.

Commented [A110]: Technical feasibility should apply to both (c) and (d)

Commented [A111]: We suggested an amendment to s25(2) in the text. However, we recommend replacing this entire subsection with the corresponding subsection from the DIS standard (s18(2)). It's unclear why there should be a difference in compliance obligations between RES and DIS services, and believe that the language in DIS s18(2) is more appropriate.

(3) Without limiting what is appropriate action, appropriate action may include exercising any of its contractual rights under the terms of use for the service in relation to the breach.

Note: For the contractual rights required to be included in terms of use see paragraph 14(1)(b).

26 Giving information about the Commissioner to end-users in Australia

(1) This section applies to the following:
- (a) a pre-assessed relevant electronic service;
- (b) a CSP~~telephony~~ relevant electronic service;
- (c) a Tier 1 relevant electronic service;
- (d) a Tier 2 relevant electronic service.

(2) The provider of a service must ensure that information:
- (a) describing the role and functions of the Commissioner; and
- (b) describing how to make a complaint to the Commissioner about the service; and
- (c) describing the mechanisms and processes required by section 27 for the service;

is accessible to end-users of the service in Australia at all times through a dedicated location on the internet site for the service.

Commented [A112]: Refer to our submission.

27 Mechanisms for end-users and account holders to report, and make complaints about, material accessible through relevant electronic services

Commented [A113]: Refer to our submission.

(1) This section applies to the following:
- (a) a pre-assessed relevant electronic service;
- (b) a CSP~~telephony~~ relevant electronic service;
- (c) a Tier 1 relevant electronic service;
- (d) a Tier 2 relevant electronic service.

(2) The provider of a service must provide a mechanism, tool or process that enables end-users in Australia to do each of the following:
- (a) identify or flag material accessible through the service as:
  - (i) in breach of an obligation under the terms of use for the service, including an obligation to comply with acceptable use policies; or
  - (ii) in breach of community standards;
- (b) report material referred to in paragraph (a) to the provider;
- (c) make a complaint to the provider about material referred to in paragraph (a).

Note: For complaints see section 40.

(3) The tool, process or technology must be available "in service", that is, not on a separate webpage.

28 Mechanisms for end-users and account holders to make complaints about breaches of this industry standard

(1) This section applies to the following services:
  (a) a pre-assessed relevant electronic service;
  (b) a CSP~~telephony~~ relevant electronic service;
  (c) a Tier 1 relevant electronic service; and
  (d) a Tier 2 relevant electronic service.

(2) The provider must provide tools, processes or technologies that enables end-users of the service in Australia to make a complaint to the provider about the provider's compliance with this industry standard.

Note: For complaints see section 40.

(3) The tools, processes or technologies must be available "in service", that is, not on a separate webpage to the webpage for the service.

29 Requirements for tools, processes and technology required under section 27 or 28 for reports and complaints

(1) A tool, process or technology required by section 27 or 28 in respect of a report or a complaint:
  (a) must be easily accessible and easy to use; and
  (b) must include or be accompanied by clear instructions on how to use them; and
  (c) must enable the person making the report or complaint to specify the harm associated with the material to which the report or complaint relates.

(2) A provider must ensure that the identity of a person who makes a report or a complaint under section 27 or 28 (the ***first end-user***) is not accessible, directly or indirectly, by any other end-user of the service without the express consent of the first end-user.

Commented [A114]: General comment about sections 27, 29 and 30 of the RES. These sections are either identical, or are close to identical to s29 of the DIS standard. From a drafting perspective, eSafety should ensure corresponding sections are consistent to simplify the standards. It's not clear why the obligations have been grouped together in the DIS, but have been split up in the RES. Recommend simplifying the RES and combining ss27, 29 and 30 into one section as it has been done in the DIS.

Commented [A115]: If a user cannot specify the harm, is then the provider at fault for not enabling the user or was the user enabled but yet unable to specify the harm. This appears unworkable and not useful. It would be more important to allow (not enable) the user to specify the material that is the subject of the complaint.

30 Appropriate steps to action reports

(1) This section applies to the following:
  (a) a pre-assessed relevant electronic service;
  (b) a CSP~~telephony~~ relevant electronic service;
  (c) a Tier 1 relevant electronic service;
  (d) a Tier 2 relevant electronic service.

(2) The provider of a service must document in writing its systems, processes or~~and~~ technologies dealing with how it responds to reports made under paragraph 27(2)(b).

(3) The provider of a service must ensure that its personnel responding to reports made under paragraph 27(2)(b) have appropriate training in and experience of the provider's policies and procedures for dealing with reports.

31 Policies and terms of use terms to be published

(1) This section applies to the following:
- (a) a pre-assessed relevant electronic service;
- (b) a CSP~~telephony~~ relevant electronic service;
- (c) a Tier 1 relevant electronic service;
- (d) a Tier 2 relevant electronic service.

(2) The provider of a service must publish:
- (a) its terms of service for the service, including its terms relating to its acceptable use policies; and
- (b) a statement setting out the community standards applicable to the service.

Commented [A116]: Refer to our submission.

(3) The publication must be accessible on the website and application (if any) for the service.

(4) The publications must:
- (a) be in plain English; and
- (b) make it clear that class 1A material is not permitted on the service and describe the broad categories of material within class 1A material; and
- (c) describe the broad categories of material within class 1B material and specify the extent to which that material is not permitted on the service, or is subject to specified restrictions.

32 Dedicated section of website for online safety information

(1) This section applies to any of the following:
- (a) a closed communication relevant electronic service;
- (b) a dating service;
- (c) a gaming service with communication functionality;
- (d) an open communication relevant electronic service;
- (e) a Tier 1 relevant electronic service.

. (2) The provider of a service must ensure that the information required by section 26 and paragraph 29(1)(b), and other online safety information made available by the provider, is accessible at all times through a dedicated location "in service", that is, not on a separate webpage to the webpage for the service.

Division 3—Reporting requirements

33 Commissioner may require risk assessments and other information

(1) The Commissioner may, by notice to the provider of a relevant electronic service, require the provider to give the Commissioner any of the following documents (if required by the standard for that service):

(a) the most recent risk profile determination for the service;

(b) the record, as required by section 10, of the most recent risk assessment for the service;

(c) the most recent assessment under paragraph 19(2)(a) for the service;

(d) the provider's development program for a specified calendar year.

Note: For development programs see section 23.

(2) The provider must give the documents to the Commissioner within the period specified in the notice.

Note 1: See also section 39.

Note 2: A provider of a relevant electronic service that is not required to prepare certain documents under this standard, such as risk assessments, will not be required to give the Commissioner such documents.

Commented [A117]: This section purports to apply to all RES, including pre-assessed RES who do not need to undertake risk assessments unless they make a material change to the service.

34 Reports of technical feasibility of compliance with provisions of Division 2

(1) The Commissioner may, by notice to the provider of a relevant electronic service, require the provider to give the Commissioner a report that specifies the extent to which it is technically feasible for the provider to comply with a specified provision of Division 2.

(2) If the report discloses that it is not, or has not been, technically feasible for the provider to use a system, process or technology as required by subsection 20(2) or (4), the report must specify the alternative action required by subsection 20(6).

Note: Section 20 is about known child sexual abuse material.

(3) If the report discloses that it is not, or has not been, technically feasible for the provider to use a system, process or technology as required by subsection 21(2) or (4), the report must specify the alternative action required by subsection 21(6).

Note: Section 21 is about known pro-terror material.

(4) The Commissioner may, by notice to the provider, require the report to be in a specified form. The provider must comply with the requirement.

(5) A report may relate to 2 or more services.

(6) The provider must give the report to the Commissioner within the period specified in the notice.

Note: See also section 39.

Commented [A118]: We believe a minimum notice period of 2 months (or 60 days) should be specified.

35 Notifying new features of relevant electronic services

(1) This section applies to the following:

(a) a closed communication relevant electronic service;
(b) a dating service;
(c) a gaming service with communication functionality;
(d) an open communication relevant electronic service;
(e) a Tier 1 relevant electronic service.

(2) If the provider of a service decides to add a new feature or function to the service, the provider must notify the Commissioner of the proposed change as soon as practicable after making the decision unless the provider considers, on reasonable grounds, that the proposed change will not ~~significantly~~ materially increase the risk that the service will be used to solicit, access, distribute or store class 1A material or class 1B material.

Commented [A119]: Use materiality concept.

(3) If a new feature or function is added to a service, the provider of the service must notify the Commissioner of the change as soon as practicable unless the provider determines, on reasonable grounds, that the change has not ~~significantly~~ materially increased the risk that the service will be used to solicit, access, distribute or store class 1A material or class 1B material.

36 Reports on outcomes of development programs

(1) The Commissioner may, by notice to the provider of a relevant electronic service to which section 23 applied in respect of a particular calendar year, require the provider to give the Commissioner a report that specifies:
(a) the activities and investments undertaken by the provider in respect of the calendar year to implement its development program; and
(b) the outcomes of those activities and investments in terms of enhancing online safety for end-users in Australia.

(2) The Commissioner may, by notice to the provider, require the report to be in a specified form. The provider must comply with the requirement.

(4) The provider must give the report to the Commissioner within the period specified in the notice.

Commented [A120]: We believe a minimum notice period of 2 months (or 60 days) should be specified.

Note: See also section 39.

37 Annual compliance reports: pre-assessed relevant electronic services and Tier 1 relevant electronic services

(1) This section applies to the following:
(a) a pre-assessed relevant electronic service;
(b) a Tier 1 relevant electronic service.

Commented [A121]: Refer to comment at table in s13 - incorrect reference in table.

(2) The provider of a service must, in accordance with this section, give the Commissioner a report (a ***compliance report***) for each calendar year during which the service was provided to end-users in Australia (each calendar year is a ***reporting period***).

(3) A compliance report under subsection (2) must include the following:

(a) the average number of monthly active users of the service in Australia during the reporting period to which the report relates, and how that number was worked out;

(b) for a Tier 1 relevant electronic service—details of the most recent risk assessment for the service, including about the plan and methodology required by subsection 9(1);

(c) for a pre-assessed relevant electronic service—a description of the service's functionalities and features during the reporting period and an explanation why the service is properly characterised as the relevant kind of service;

(d) details of the steps that the provider took during the reporting period to comply with the requirements of this Part;

(e) an explanation why the steps taken as mentioned in paragraph (d) were appropriate, having regard, among other things, to the features of the service during the reporting period;

(f) a statement of the extent to which it was not, during the reporting period, technically feasible for the provider to detect or remove class 1A material or class 1B material from the service, and why;

(g) where it was technically feasible, the amount of child sexual exploitation material and pro-terror material that the provider removed from the service during the reporting period;

(h) where it was technically feasible, details of how the child sexual exploitation material and pro-terror material that the provider removed from the service during the reporting period was detected and identified;

(i) the number of complaints made to the provider about the provider's compliance with this industry standard during the reporting period.

Note: For paragraph (g), examples include end-user reports and use of hashing technologies.

(4) The report must provide justification for the conclusions in the report.

(5) The Commissioner may, by notice to the provider, require the compliance report to be in a specified form. The provider must comply with the requirement.

(6) A compliance report may relate to 2 or more services.

(7) If information required to be included in a compliance report has otherwise been given to the Commissioner, the provider may refer to the report or notification by which it was given instead of repeating it in the compliance report.

(8) A compliance report must be given to the Commissioner within 2 months after the end of the reporting period.

Note: See also section 39.

38 Compliance and other certificates and reports required by Commissioner

Enterprise relevant electronic services

(1) The Commissioner may, by notice to the provider of an enterprise relevant electronic service, require the provider to certify that~~, except as specified in the~~

~~certificate,~~ the provider has complied with section 14 during the immediate past calendar year or, in the case of non-compliance, indicate areas of non-compliance.

Other relevant electronic services

(2) The Commissioner may, by notice to the provider of any of the following:
(a) a closed communication relevant electronic service;
(b) a dating service;
(c) a gaming service with communications functionality;
(d) a CSP~~telephony~~ relevant electronic service;
(e) a Tier 2 relevant electronic service;
require the provider to give the Commissioner a report (in this section, a ***compliance report***) for the immediately preceding calendar year (in this section, the ***reporting period***).

(3) A compliance report under subsection (2) must include the following:
(a) for a CSP~~telephony~~ relevant electronic service—a description of the service's functionalities and features and an explanation why the service is properly characterised as a CSP~~telephony~~ relevant electronic service;
(b) for a Tier 2 relevant electronic service—details of the most recent risk assessment, including about the plan and methodology required by subsection 9(1);
(c) in any case:
(i) details of the steps that the provider took during the reporting period to comply with the requirements of Part 4;
(ii) an explanation why the steps taken as mentioned in subparagraph (c)(i) were appropriate, having regard, among other things, to the features of the service during the reporting period;
(f) a statement of the extent to which it was not, during the reporting period, technically feasible for the provider to detect or remove class 1A material or class 1B material from the service, and why.

(4) A compliance report under subsection (3) must provide justification for the conclusions in the report.

(5) The Commissioner may, by notice to the provider, require the compliance report to be in a specified form. The provider must comply with the requirement.

(6) A compliance report may relate to 2 or more services.

Giving certificates and reports

(7) A provider must comply with a notice under this section within 2 months after service of the notice on the provider.

Note: See also section 39.

Commented [A122]: The drafting is unclear - is this what is meant?

Commented [A123]: These services are pre-assessed and therefore already captured under s37.

Commented [A124]: Refer to comment at table in s13 - incorrect reference in table.

39 Extension of reporting periods

The Commissioner may, on application, extend the period within which a provider must give the Commissioner a report, certificate or notification under this Division, and may do so before or after the period has expired.

Part 5—Miscellaneous

40 Complaint resolution arrangements

(1) This section applies to a relevant electronic service if this industry standard requires the provider to make provision in respect of complaints by end-users in Australia of the service.

(2) If a complaint in relation to the service is made by an end-user, the provider must:
   (a) investigate the complaint; and
   (b) notify the complainant of the outcome of the investigations and the action proposed by the provider to in consequence of the investigation.

(3) Subsection (2) does not apply if:
   (a) the provider believes on reasonable grounds that the complaint was frivolous or vexatious or otherwise not made in good faith; or
   (b) the matter the subject of the complaint is being investigated, or has been investigated, by the Commissioner under Division 5 of Part 3 of the Act.

Commented [A125]: Refer to our submission in relation to providers who do not have visibility or access of the content.

41 Record-keeping requirements

(1) The section applies to all relevant electronic services.

(2) The provider of a service must keep records that set out the actions that the provider has taken to comply with this industry standard.

(3) The provider must keep the records for at least 2 years after the end of the calendar year during which the action was taken.

Published by:
COMMUNICATIONS
ALLIANCE LTD

Level 12
75 Miller Street
North Sydney
NSW 2060 Australia

Correspondence
PO Box 444
Milsons Point
NSW 1565

T 61 2 9959 9111
F 61 2 9954 6136
E
info@commsalliance.com.au
www.commsalliance.com.au
ABN 56 078 026 507