To: Ms Julie Inman-Grant, Office of the eSafety Commissioner
By email: submissions@esafety.gov.au

Monday 22 January, 2024

Dear Ms Inman-Grant,

The Digital Industry Group Inc (DIGI) thanks you for the opportunity to provide our views on the establishment of two new industry Standards under the Online Safety Act 2021 (OSA): Online Safety (Relevant Electronic Services – class 1A and 1B Material) Industry Standard 2024 (the Relevant Electronic Services Standard) and the Online Safety (Designated Internet Services Standard – class 1A and 1B Material) Industry Standard 2024 (the Designated Internet Services Standard).

DIGI is a non-profit industry association that advocates for the interests of the digital industry in Australia. DIGI's founding members are Apple, Discord, eBay, Google, Linktree, Meta, Microsoft, TikTok, Twitch, Spotify, Snap, Yahoo, and X. DIGI's vision is a thriving Australian digitally-enabled economy that fosters innovation, a growing selection of digital products and services, and where online safety and privacy are protected.

DIGI wishes to underscore that our members share and support the Government's commitment to combating the risks posed by seriously harmful online materials. DIGI, alongside Communications Alliance, played a leading role between 2021 and 2023 in the drafting of the eight codes of practice under the OSA (OSA Codes) to protect Australians from class 1A and 1B materials, being categories of materials that would be refused classification under Australia's classification scheme. As a result of our role in the Code development process, DIGI has unique insights concerning the technical challenges of regulating class 1A and class 1B materials for relevant electronic services and designated internet services which have informed our submission.

The submission is divided into 4 parts:

- Part 1: Summary of issues and recommendations;
- Part 2: Common issues relevant to both DIS and RES Standards;
- Part 3: Issues specific to the RES Standard; and
- Part 4: Issues specific to the DIS Standard.

A core concern of DIGI, throughout this submission, is the considerable variance between the approach and scope of the OSA Codes and the draft Standards. These variances create various compliance challenges for industry and the rationale for many of the variances is unclear. As eSafety has acknowledged, there is a degree of overlap between three sections of the online industry: social media services, relevant electronic services and designated internet services. To the extent that service providers undertake similar activities, it is reasonable for industry and users to expect that they will be regulated in broadly the same way. We are concerned that, unless addressed, the variances between the OSA Codes and the Standards will create two inconsistent regimes for regulation of class 1 materials in Australia, making it hard for users to understand their rights and for industry to comply with their obligations across the range of services regulated by Part 9 of the OSA.

Our submission does not address all the questions set out in the Discussion paper developed for this consultation.[1] eSafety has extensively discussed many of these with DIGI and other industry participants over the course of the OSA Codes development process. DIGI's submission therefore touches on some, but not all, of the questions raised by eSafety and, additionally, raises a range of issues that have not been canvassed in the Discussion paper. DIGI has identified a range of issues that we consider should be addressed by eSafety in its written guidance and recommend that industry has an opportunity to provide feedback on the guidance before it is finalised. We have also made specific suggestions and recommendations for changes to the Standards, which we hope will be helpful in ensuring that the Standards for relevant electronic services and designated internet services provide appropriate community safeguards for end-users in Australia in relation to class 1A and class 1B materials.

We also note that, in parallel with this consultation, the Government is seeking public feedback on a proposal to make changes to the Online Safety (Basic Online Safety Expectations) Determination 2022.[2] DIGI will make a separate submission on those proposed changes, but we note that there are several areas of overlap between the proposed standards and the BOSE. While the BOSE, Codes and Standards are separate instruments under the OSA, they share a common purpose of enhancing online safety. With that purpose in mind, we think it is important to ensure that the requirements of the Standards, Codes and BOSE set a consistent standard for compliance to the extent they deal with class 1 materials and related subject matter. We are concerned that, as drafted, the proposed changes to the BOSE contain several duplicative and potentially inconsistent requirements on industry for the same subject-matter covered by the Codes and Standards, including requirements around enforcement of terms and conditions, addressing the risk of generative AI, and transparency reporting. For example, it is unclear whether a participant that meets requirements in the Codes regarding enforcement of terms and conditions for class 1 materials, would necessarily satisfy the same requirements for enforcement of terms and conditions under the BOSE. These variances need to be addressed so that industry participants are clear about the steps they need to take to fulfil their compliance obligations under the OSA.

We thank you for your consideration of the matters raised in this submission. DIGI considers itself a key partner in the Government's efforts to address online safety, and we look forward to our continued dialogue on our shared goals in this area. Should you have any questions about this submission, please do not hesitate to contact me.

Yours sincerely,

Jennifer Duxbury

Dr Jennifer Duxbury
Director Policy, Regulatory Affairs and Research
Digital Industry Group Inc. (DIGI)

[1] eSafety Commissioner, *Discussion paper draft Online Safety (Relevant Electronic Services – class 1A and 1B Material). Industry Standard 2024 and draft Online. Safety (Designated Internet Services Standard – class 1A and 1B Material) Industry. Standard 2024* (November 2021).

[2] *Amending the Online Safety (Basic Online Safety Expectations) Determination 2022: Consultation Paper* (November 2023)

Part 1: Summary of Issues and Recommendations

DIGI makes **48 specific recommendations** in this submission, drawing on our experience developing six of the OSA Codes that were registered by the eSafety Commissioner. This section of the submission contains a summary of DIGI's specific suggestions regarding amendments to the Standards and accompanying guidance. In addition, as a general observation, we note that, unlike the OSA Codes, guidance on the mandatory compliance measures in the draft Standards is not embedded in the regulations but will rather be contained in other written regulatory guidance developed by eSafety subsequent to this consultation. Without visibility of this guidance, it is difficult for industry to interpret eSafety's expectations on how certain requirements are to be implemented in practice.

Common Issues relevant to both DIS and RES Standards

Guidance

A. Given the critical importance of eSafety's written guidance for the Standards, we would appreciate it if eSafety could provide an opportunity for industry to comment on the guidance before it is finalised.

Definitions and terminology

B. We suggest that section 5 of the RES and DIS Standards be amended so that industry participants must apply the same 'predominant purpose' test for determining which Code or Standard is applicable to its online activities in Australia.

C. The definitions of class 1A and class 1B materials should reference all the relevant requirements of the classification scheme as set out in a Schedule.

D. The definitions of child sexual abuse material or CSAM, child sexual exploitation material or CSEM, crime and violence material, drug-related material, extreme crime and violence material, and pro-terror material should all operate as subsets of class 1 material (as defined in the OSA) and directly incorporate the approach to classifying such content specified by the classification scheme, rather than copy pasting selected parts of the Classification Act that may change.

E. The Standards should make it clear that industry participants may use different terminology to describe class 1A material, class 1B material, and the subcategories of class 1A material and class 1B materials for different audiences.

Classification of class 1 materials

F. We suggest that the Standards, or at a minimum the guidance for the Standards, should follow the OSA Codes and provide that:

Industry participants are not required to arrange for any material to be classified or otherwise to replicate the classification functions of the classification Board. However, where an industry participant is required by a Standard to identify or distinguish between different sub-categories of unclassified class 1 material, the industry participant must develop a process for categorising that material in a way that is informed by the classification process. In each case, the process developed by the industry participant will serve as a proxy for the classification process and may vary depending on the circumstances in which it is to be applied, including to take into account the nature and volume of material to be categorised. Where the Standard is to be applied in relation to a single known item of material, the process may involve a detailed review of that material and may follow the classification process more closely than where the compliance measure is to be applied at scale in relation to multiple unknown items of material. For example, when applying a Standard at scale, an industry participant may develop particular flags or triggers in order to identify material that is likely to be class 1A or class 1B material under this Code, where those flags or triggers are framed by reference to factors that would be taken into account pursuant to the classification process. Different industry participants may develop different flags or triggers, depending on the nature of the online activities they are undertaking, the nature of the material they are dealing with, and other relevant factors.

G. We suggest that the Standards make it clear that industry participants will be treated in an equivalent way as provided for in the Head Terms i.e that eSafety will not hold a participant in breach of a Standard because they categorise material differently to other participants or to eSafety.

Account-holder concept

H. We recommend that eSafety reconsider the need for this concept given the proposed approach does not appear to achieve any workable distinction and it is unclear what it is seeking to achieve.

'Appropriate' as a standard of compliance

I. We suggest that sections 12 of the Standards be amended to provide as follows:

Where the term appropriate is used to describe the actions required by the provider of a service to comply with a standard, a provider is required to demonstrate that its actions are reasonable in the circumstances, taking into account all relevant considerations including:

a) the extent to which the action is likely *to* achieve 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 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.

J. The Standards should provide that the factors listed in section 5.1 (b) of the Head Terms of the OSA Codes are also relevant considerations in determining whether action is appropriate.

Approach to proactive detection, technical feasibility and encryption

K. The Standards should provide greater flexibility for providers of services to best tackle the risks of CSAM and pro-terror materials in a manner that is appropriate for individual services given the broad range of service types covered by these Standards, and the complex privacy and security considerations that accompany many of these services.

To the extent that provisions contain proactive detection obligations for CSAM, recommendations L to R are also made.

L. The Standards should be more broadly subject to the technical feasibility test. For instance, all detection obligations, obligations contingent on visibility of content/communications and other obligations requiring technological solutions in the Standards (including requirements to take appropriate alternative action) should be subject to the technical feasibility test. For example, the proactive detection provisions in the DIS Standard relating to end-user hosting services should be subject to the technical feasibility test and providers should not be required to take 'appropriate alternative action' that is not technically feasible.

M. The proactive detection requirement should allow industry participants the flexibility to implement systems, processes or technologies that are designed to detect, flag and/or remove from the service, instances of known CSAM, to ensure a consistent approach to this measure across RES, DIS, and SMS services.

N. The Standards should be amended to make clear that they do not require any industry participant to undertake steps that do the following (regardless of the costs involved):

a) implement or build a systematic weakness, or a systematic vulnerability, into a form of encrypted service or other information security measure such as private firewall configurations, VPN tunnels and private networking links, which work directly or complement encryption to protect legitimate cybersecurity and data integrity interests;

b) build a new decryption capability in relation to encrypted services;

c) render methods of encryption less effective; or

d) undertake monitoring of private communications between end-users.

O. The considerations that should be taken into account in applying the technical feasibility test in sections 7 of the Standards should include new considerations: '(c) whether it is reasonably practical for the provider to review or remove material on the service' and '(d) the accuracy and reliability of the available proactive detection systems, processed or technologies'. In addition, the Standards should be revised to clarify that cost is only one consideration for technical feasibility (i.e., lower cost is not viewed as a controlling factor, by itself, that an action is feasible). The Standards should reflect the fact that technical feasibility encompasses a range of concerns going to feasibility – not just whether something is technically possible or feasible, but also whether it is (i) reasonable taking into account concerns such as security and privacy risks (e.g., whether it would negatively impact a provider's security architecture, design or operating system); and (ii) feasible from a legal and practical perspective (see comments on section 17 of this submission below).

P. We suggest that the additional requirements for end-user managed hosting services and Tier 1 (see section 21(8)) are unnecessary given the remainder of the measure.

Q. Clarity should be included in the Standard regarding what 'appropriate alternative action' could be for fully end-to-end encrypted services.

R. The guidance for compliance with the proactive detection provisions concerning CSAM should direct providers to consider:

a) *the availability of different options for proactive detection of known CSAM and the capability of the provider to use those options accurately;*
b) *the need for human resourcing required to review detected materials, for example, where an image that has been detected has been altered to make it more difficult for it to be matched with the original image;*
c) *the need to provide adequate health and safety arrangements for personnel undertaking the review of known CSAM; and*
d) *the fact that hash lists of known CSAM are not infallible. Providers should therefore take care to safeguard against low quality hashes and hashes prone to collisions (e.g . compilation videos) by having a suitable confirmation and quality control process to independently confirm that the material depicted in the hash is CSAM. Where a hash is likely to lead to false results, a provider should not deploy it.*

Proactive Detection of pro-terror materials

Please see Recommendation K, which also applies to pro-terror materials. To the extent that the Standards contain provisions relating to proactive detection of pro-terror materials, the following Recommendations are also made:

S. Recommendations L, M, N, O, Q and R should also apply to pro-terror materials.

T. For consistency with the definitions of pro-terror material in the classification scheme, the Standards should be amended to make clear that the requirements to proactively detect pro-terror materials apply only to the extent that the material has been sent to or shared with another person via the relevant DIS or RES service and not to material that is stored on the service in an 'inert state'.

U. The Standards should limit the requirement to proactively detect known pro-terror material to relevant RES and DIS services that have more than 2 million active Australian end-users so as to not impose an unreasonable burden on small to medium size businesses who may not be capable of complying with the requirements.

V. The Standards should be accompanied by guidance that makes clear that providers need to ensure detected material is subject to human review before it is classified a pro-terror material to ensure the context of the material is considered in accordance with the requirements of the classification scheme. In addition similar guidance needs to be provided as for known CSAM concerning the compliance systems that need to be in place e.g to prioritise material for human review and protect the health and safety of personnel that are reviewing potential pro-terror material.

W. The absence of any organisation that can verify whether material is pro-terror material or the definition of pro-terror material is a significant issue for industry. As drafted it is not clear how eSafety expects industry to access 'verified material', which is of real concern as different nation-states and international bodies, such as the UN, take widely different approaches to determining what material is terrorist in nature. In the absence of any organisation that can verify whether material is pro-terror material, the definition of pro-terror material should be amended to make clear that 'known-pro-terror material' includes material that eSafety has previously verified as pro-terror material consistent with the approach of the OSA Codes. Further consideration should be given to establishing systems and processes within eSafety for verifying pro-terror material, including in cooperation with the Classification Board, and making its decisions about pro-terror material available to industry.

Requirements to report materials to verifying authorities

X. We think that the obligations to notify a verifying authority for CSAM and pro-terror materials should be removed given there are no authorities that maintain databases of verified CSAM or pro-terror materials under Australian law. Alternatively the obligation should be limited to notifying materials that a provider reasonably believes are not known CSAM to a non-governmental expert child protection authority which maintains a database of materials that may be accessed by industry participants subject to the Standards, in order to proactively detect CSAM on their services.

Disrupt and deter requirements

Please see Recommendation K, which also applies to the obligations to disrupt or deter CSAM and pro-terror material. To the extent the Standards contain disrupt and deter provisions, the following Recommendations are also made:

Y. We suggest that these disrupt and deter requirements in the Standards for CSAM and pro-terror material be amended so that:

a) The measure: (i) provides more flexibility for providers to achieve compliance in a number of ways – given the range of providers covered and the early stage of relevant technologies; and (ii) ensures that industry commits to continuously improving efforts in a way that is proportionate to the risk without requiring the implementation of effective tools as of day one (which is an unrealistic bar, given that many of these technologies are still early stage for a range of service offerings and not yet effective).

b) The requirement to implement 'effective systems, processes and technologies that disrupt and deter CSAM and pro-terror material' is replaced by a requirement by reference to 'appropriate systems, processes and/or technologies that disrupt and deter CSAM and pro-terror material'.[3]

c) If the requirement to implement technology is retained, it should be limited to providers with more than 2 million active end-users in circumstances where there is a reasonable suspicion there is new generation (un-known) CSAM or pro-terror material on the service.

d) The technical feasibility test should apply to this measure, including as modified by L, N and O above (i.e. providers are not required to take any action that is not technically feasible);

e) Enterprise services should be excluded from the disrupt and deter pro-terror requirements; and

[3] See discussion on need to revise the meaning of 'appropriate' in section 5 of this submission

f) The requirements to disrupt and deter pro-terror material should not apply to the relevant DIS or RES services to the extent that material is stored on the service in an 'inert state', and is not accessible to another person.

Z. eSafety should also provide detailed guidance that explains the risk of 'false positives' and how these risks should be mitigated by providers, including guidance concerning the need for the provider to make suitable arrangements for the human review of detected materials and the need to accurately verify the context of the detected materials before enforcement action is taken under the provider's terms of use.

Development programs

AA. We suggest that the requirements for providers to establish a development program should be removed.

BB. If retained, at a minimum they should be amended as follows:

a) section 24 (3) of the DIS Standard and section 23 (3) of the RES Standard should be amended to provide that 'the provider of the service must establish and implement, for the calendar year, a program of development (***development program***) in respect of systems, processes or technologies that is *reasonably proportionate to the risk of class 1 material on the service* as assessed in accordance with this Standard;

b) the mandatory requirements should be discretionary i.e., the word 'must' in section 23 (4) of RES Standard and section 24 (4) of the DIS standard should be replaced by 'may'; and

c) the requirements of this obligation should only apply to services with more than 2 million active monthly end-users.

CC. Section 37 of the DIS Standard and 36 of the RES Standard should be deleted. We suggest that the requirements for providers to establish a development program should be removed. If retained, at a minimum they should be amended as follows:

a) the primary requirement in section 24 (3) of the DIS and section 23 (3) of the RES is redrafted to provide that 'the provider of the service must establish and implement, for the calendar year, a program of development (***development program***) in respect of systems, processes *or* technologies that is *reasonably proportionate to the risk of class 1A or class 1A material on the service* as assessed in accordance with this standard;

b) the mandatory requirements should be discretionary i.e., the word 'must' in section 23 (4) of RES Standard and section 24 (4) of the DIS standard should be replaced by 'may'; and

c) The requirements of these sections should only apply to services with more than 2 million active monthly end-users.

Terms of Use

DD. We suggest that the requirements in sections 14 be amended as set out below:

a) The primary requirement in sections 14 (2) should provide that 'the provider of a service must include provisions in the terms of use for the service that:

i) prohibit end-users from using the service for CSEM or pro-terror material; and

ii) contain appropriate restrictions on end-users concerning other categories of class 1 materials.'

b) The inconsistencies in the treatment of class 1A material other than CSEM and pro-terror material, and of class 1B material, between the RES and DIS Standards should be rectified.

c) The requirement in section 14(2)(b) should be limited to CSAM and pro-terror materials.

d) The requirement on providers of enterprise services in section 14(2)(b)(ii) to include provisions that enable them to suspend the provision of the service to a specified end-user of the service for a specified period if they breach the prohibition on class 1 materials should be removed.

e) Section 14(3) should be amended as follows:

> *Where technically feasible, where 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 reasonably proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach.*

The changes to the technical feasibility test recommended in part 6 of this submission above should also apply here. Please also see recommendations in part 17 of this submission below.

f) The changes to the evidentiary burden in sections 14(4) should be removed.

g) That the additional requirement for end-user managed hosting services in section 16(6) of the DIS Standard is either removed (as unnecessary) or that clarity is provided as to what else is required by section 16(6) in addition to what is already covered by section 14.

Transition period for compliance

EE. The Standards be amended to make allowance for an additional 6 month period for services to be fully compliant, consistent with section 7.2(b)(iii) of the Head Terms of the OSA Codes.

Risk assessments following changes to a service

FF. We suggest that prohibition on changes to a service that increase risk for class 1A and 1B materials (other than CSEM) apply only where the change materially increases the risk that such materials will cause harm to other end-users of the service in Australia. The Standards, or accompanying guidance, should also make clear that employing new or different encryption or other information security measures shall not be deemed to increase the risk for class 1 materials.

GG. We suggest that the Standards be amended so that services that have pre-allocated compliance requirements under the Standards are only required to carry out risk assessments when they make a material change in the service that changes the way the service is categorised under the Standard (and thus the requirements that apply to the service).

Community Standards

HH. The term ‘community standards’ should be removed to avoid confusion if it is intended to be equivalent to terms and conditions and acceptable use policies (as per the OSA Codes). The term ‘community standards’ be replaced by another clearly defined term in the Standards, if this term is intended to cover something different to terms and conditions and acceptable use policies.

Engagement with user reports of class 1 materials

II. We suggest that the Standards be amended so that it is clear the requirements to respond to reports and take specific enforcement action, including removal of material, only apply where the provider is aware that the specific instances of the material is in breach of its terms and conditions and, in relation to materials that are the subject of a report, that it is technically feasible for the provider to investigate and independently verify the accuracy of the report. See also: discussion and recommendations on technical feasibility in section 6 of this submission.

JJ. See recommendations in section 11 of this submission regarding the range of enforcement options that a provider may consider. As recommended in section 11, this should not be prescriptive and a provider should instead be required to enforce its contractual rights in respect of the breach in an appropriate way that is reasonably proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach. Requirements to take enforcement action should be consistent with the action required by section 14 (subject to our recommendations in section 11 of this submission) and should be subject to the technical feasibility test.

KK. The requirements to achieve removal of all instances of material should be deleted.

Requirements to provide information ‘in-service”

LL. We suggest that the Standards be amended so that the obligation is to provide users with certain information/tools in a manner that is ‘that is clear and accessible to end-users of the service in Australia’.

Annual reporting requirements

MM. We suggest that reports be confined to:

a) the steps that the provider has taken, including measures and controls the provider has implemented, to comply with applicable minimum compliance measures in the Standard;
b) confirmation from the provider that the steps, measures and controls are appropriate, including reasonable supporting details and evidence;
c) where applicable, details of any limitations in terms of technical feasibility on the service or the provider to identify, assess or take action in respect of class 1A material and class 1B material as required by the Standard;

d) where applicable, details of the most recent risk assessment for the service, including about the plan and methodology; and

e) details of the volume of CSEM and pro-terror materials removed from the service.

NN. The reporting requirements should be amended to ensure that industry participants are given a minimum of 2 months notice to comply with a request for a report, consistent with the equivalent provisions in the OSA codes.

Complaints handling

OO. We suggest that sections 40(2)(b) of the Standards be amended to require providers to notify complainants only of the decision and action the providers have taken in respect of the information to which the complaint relates, where the complainant has provided contact information.

Issues specific to RES standard

Duplicative risk assessment requirements for pre-assessed relevant electronic services

PP. The references in section 38(2) to 'closed communication relevant electronic service; a dating service; and a gaming service with communications functionality' should be removed as these services reporting obligations are already specified in section 37.

Primary functionality as test for distinguishing RES categories

QQ. We suggest that the tests of 'primary functionality' for the different categories of RES services be removed as it is confusing and difficult to apply.

Open communication relevant electronic services

RR. We suggest that the avoidance of 'doubt provision' in the definition of open communication service should be removed as it is confusing.

Tier 3 RES

SS. We suggest that the Standard be amended to include a definition of a Tier 3 RES.

Issues specific to DIS standard

Operating systems should not be in scope of the standard

TT. We suggest that the Standard be amended to make clear that operating systems, or to OS providers (to the extent providers are acting in that capacity), as defined in the Equipment Code are not subject to the Standard.

Classified DIS

UU. We suggest that the definition of classified DIS make explicit reference to the relevant categories under the Classification Guidelines for Publications.

High impact generative AI DIS and machine learning model platforms

VV. We suggest that eSafety remove the requirements for high impact generative AI DIS and machine learning platforms for the time being. We consider that the requirements for the regulation of these services will benefit from a more extended public consultation on how to work through the complex issues raised by the proposed requirements for these service types. This may best be achieved as part of the review of the Online Safety Act 2021, which is scheduled to commence early in 2024.[4]

Part 2: Common issues relevant to both DIS and RES Standards

The proposed Standards in relation to relevant electronic services and designated internet services that are the subject of this consultation have been prepared by the eSafety Commissioner following the Commissioner's decision that the two codes for these service types proposed by the industry did not, in the Commissioner's view, meet the statutory requirements for registration. The basis for the eSafety Commissioner's decision is set out in the Statements of Reasons provided on 9 June 2023. Whilst it is clear that the draft Standards have drawn upon the draft OSA Codes developed by industry for RES and DIS providers they differ significantly from the draft and registered OSA Codes on many levels. There are key differences between the standard of compliance required by industry participants under the Standards and the Codes. For example, the Standards do not carry across important provisions in the Head Terms of the OSA Codes that are designed to ensure that industry participants take a reasonably proportionate approach to implementing compliance measures based on a range of considerations.

1. Conflicting tests for determining the applicability of Standards versus Codes: RES Standard, section 5 and DIS Standard, section 5

The Standards take a different approach to the Codes in determining whether a Standard applies to services of a particular industry participant. We are concerned that this creates confusion and

[4] Michelle Rowland MP, *Media Release*, 22 November 2023 accessible at https://minister.infrastructure.gov.au/rowland/media-release/albanese-government-takes-major-steps-forward-improve-online-safety.

uncertainty for industry participants in determining which regulatory instrument applies to their services .

1.1. **Industry's approach:** The Head Terms of the OSA Codes sets out the test that industry participants must apply in determining which Code or Standard is applicable to its online activities in Australia:

Where a single electronic service could fall within the scope of more than one industry code or industry Standard, the relevant industry participant will only be required to comply with one code or industry Standard, as the case may be, for that electronic service. The code or industry Standard that will apply in this situation is the code or industry Standard that is most closely aligned with the ***predominant purpose*** *of the single electronic service*[5].

We agree that participants should only be required to comply with one Code or Standard per service. However, the ***predominant purpose of a service*** is in our view the most suitable test for determining the applicability of an OSA Code or Standard, because many services have a range of similar functionalities.For example, many SMS, RES and DIS enable users to post and download content, communicate with other users, and store material online[6]. The purposive approach was adopted by industry in the OSA Codes and is consistent with the approach adopted by the OSA. So for example, in order to be a social media service as defined under the OSA the sole or primary purpose of the service must be 'to enable online social interaction between 2 or more end-users'.

1.2. **eSafety's revised approach:** Question 3 of the Discussion paper asks whether the test in section 5 workable? Is further guidance required to assist providers to determine whether this Standard, or another code or Standard, applies to a particular online service? In our view the test is confusing and not workable as drafted. The predominant functionality test in section 5 is in direct conflict with the equivalent test in the Codes and is potentially inconsistent with the test used in OSA (to the extent there is one, noting that 'purpose' is part of the definition of social media services). eSafety has not explained why this different approach was adopted or how industry participants should reconcile these different tests.This issue is further exacerbated by the way eSafety has defined subcategories of services by reference to their functionality. For example, under eSafety's proposed approach a relevant electronic service is a closed communication relevant electronic service if its 'primary functionality' includes enabling end-users to 'create a list of other end-users of the service' and accessing and communicating with people on that list.[7] However, the ability to 'create a list' is not the *primary functionality* of most relevant electronic services – it is an ancillary function. For instance, a user of an email service

[5] See *Consolidated Industry Codes of Practice for the Online Industry (class 1A and class 1B Material) Head Terms (Head Terms)*, Preamble page 4.

[6] These different functionalities are, however, acknowledged in the *Head Terms* to be critical to the design of the measures in the Codes, Preamble, Page 5: *"Electronic products and services provided across different sections of the online industry may include different functionalities, which in turn may be relevant to: • the connection between a product or service and a relevant online safety risk; • the relationship between a product or service and an end-user, including whether or not a service controls the end-user interface; and • the visibility, control or administration of specific material accessible to an end-users.*

[7] See further, section 22 of this submission.

can choose to email other individuals without adding them to their contact list. It is the *enabling of communication that is both the primary purpose and function of such services.*

1.3. **Suggested Revisions:** The emphasis on the functionality of services throughout the Standards creates significant confusion for industry in determining which Code or Standard is relevant to their services. We suggest this is best addressed by adopting the test used in the OSA Codes (given that eSafety did not raise concerns about the purposive test in the Head Terms in the OSA Codes' development process).

Specific recommendation

B. We suggest that section 5 of the RES and DIS Standards be amended so that industry participants must apply the same 'predominant purpose' test for determining which Code or Standard is applicable to its online activities in Australia.

2. Different definitions and approaches to defining subcategories of class 1 Materials in Codes and Standards is confusing for industry participants and users

The sub-categories of materials that are regulated under Part 9 of the OSA, are defined very differently in the draft Standards and the Codes. While the industry Codes seek to adopt the approach of the OSA and the classification scheme to defining these subcategories, eSafety has modified the approach of the scheme. Rather than referring to relevant definitions in the OSA and/or the classification scheme, the Standards include new standalone definitions of various subcategories of class 1 materials. As a result, there is inconsistency between the Standards and the OSA, and a risk that the Standards extend to content beyond class 1 material. The inconsistency between the way subcategories of class 1 materials are defined in the draft Standards and the OSA is also confusing and creates a range of compliance challenges for industry participants.

2.1. **Approach of Codes to defining subcategories of class 1A and 1B materials:** The Head Terms of the OSA Codes were designed to ensure that as far as possible, the definition of class 1 materials and the subcategories of materials that fall within class 1A and class 1B are consistent with the OSA. Section 106 of the OSA makes clear that class 1 material is that which has or would be classified as RC under the *Classification (Publications, Films and Computer Games) Act 1995* (the Classification Act). So, for example, the Head Terms define the two subcategories of materials that are covered by the Codes as follows:

> *Class 1A material is a sub-category of class 1 material used for the purpose of this Code that comprises child sexual exploitation material, pro-terror material, and*

> *extreme crime and violence material, in each case as further described in Annexure A.*
>
> *Class 1B material is a sub-category of class 1 material used for the purpose of this Code that is composed of crime and violence material and drug-related material, in each case as further described in Annexure A.*
>
> *Annexure A provides an overview of the classification process as it applies at the date of this Code. To the extent that the classification process changes, including as a result of any legislative reforms or changes to applicable supporting codes or guidelines, industry participants must refer to the updated classification process. The industry representatives responsible for leading the development of this Code must work to promptly update this Annexure as required to reflect any changes to the classification process.*

2.2. The approach of the OSA Codes was intended to ensure that any updates to the classification scheme (currently under review) that are relevant to the regulated subcategories of class 1 materials are incorporated into the Codes. Critically the OSA Codes are drafted so as to ensure that industry is aware of the relevance of the full range of legislation that forms the National Classification Scheme (classification scheme) and which needs to be considered in classifying content such as the Classification Act and Classification Guidelines including the essential principles that must be applied by classifiers such as 'the crucial importance of context'.[8]

2.3. Furthermore, the industry approach in the OSA Codes acknowledges that the terminology of the classification scheme may be unfamiliar to users and that when implementing certain measures, it is desirable to enable industry to use more user-friendly terminology e.g., in their terms of use and policies.[9] In contrast the Standards imply that industry must strictly use technical terms such as 'class 1 materials' that will be unfamiliar to users and which will make it difficult for them to identify and report relevant online material to providers.

2.4. **Approach of DIS and RES Standards to define subcategories of class 1A and IB materials:** the definitions of sub-categories of materials in the Standards simplify how the classification scheme currently defines the relevant types of materials that are refused classification offline by creating new standalone definitions. This approach fixes the meaning of the sub-categories of class 1 materials at the date the Standard is implemented in a way that is detached from the requirements of the classification scheme.[10] In effect, the Standards appear to amend the more nuanced approach to classifying material required under the classification scheme. This creates several issues. Firstly, as soon as the classification scheme is amended in any significant way, the entire Standard will require amendment and will operate in an inconsistent manner until that occurs. Secondly, as drafted, the definitions of the sub-categories of materials omit reference to key requirements of the classification scheme. For example, the

[8] Guidelines for the classification of Computer Games 2012, Guidelines for the Classification of Films 2012 and Guidelines for the Classification of Publications 2005.

[9] Sec 3 (e), *Head Terms*.

[10] It is worth noting that the National Classification Scheme is currently under review, and therefore changes to the Scheme are likely.

principles underpinning the National Classification Code as well as the principles underpinning various Guidelines (e.g. the Film Guidelines) are entirely omitted. The principles underpinning the Film Guidelines (by way of example) include detail about the principles that should underlie the use of the Guidelines, such as the importance of context, the way to assess the impact of material, and how to apply the classifiable elements. By seeking to paraphrase the approach of the OSA and classification scheme into standalone definitions, the requirements of the OSA and classification scheme have been amended (and in some instances entirely omitted) resulting in inconsistency with the definition of class 1 material in the OSA, as well as the approach of the OSA Codes, which specifically directs industry participants to the relevant classification scheme requirements.

2.5. **Suggested revisions:** Clearly, from a compliance and enforcement perspective, it is desirable that both the OSA Codes and Standards define class 1 material and sub-categories of class 1 materials in the same way, in the manner required by the OSA and the classification scheme to ensure that the Standards relate only to class 1 material as defined in the OSA. Further, it would be helpful to be clear that industry may use different terminology to describe the categories and sub-categories of materials, where appropriate.

Specific recommendations

C. The definitions of class 1A and class 1B materials should reference all the relevant requirements of the classification scheme as set out in a Schedule.

D. The definitions of child sexual abuse material or CSAM, child sexual exploitation material or CSEM, crime and violence material, drug-related material, extreme crime and violence material, and pro-terror material should all operate as sub-sets of class 1 material (as defined in the OSA) and directly incorporate the approach to classifying such content specified by the classification scheme, rather than copy pasting selected parts of the Classification Act that may change.

E. The Standards should make it clear that industry participants may use different terminology to describe class 1A material, class 1B material and the sub-categories of class 1A material and class 1B materials for different audiences.

3. Need for guidance for industry on how providers should deal with classifying online materials

The classification scheme was not developed to regulate vast amounts of online materials at scale and only applies to discrete types of materials, before commercial release (largely for entertainment purposes). Consequently, there is considerable uncertainty regarding how industry should practically go about classifying online materials in accordance with the scheme. For this reason, the OSA Codes (section 3(f) of the Head Terms) provided that industry participants were not expected to replicate the processes used to classify offline material, but could develop their own processes. In addition, section 3(g) and section 5.3 of the Head Terms, provided that the fact that a provider categorised a given item of content in a different way to another or to eSafety, or had not identified every item of class 1A or class 1B material on its service did not indicate a failure to comply (provided of course that its processes had been developed and applied reasonably). This reflects the challenges of trying to apply concepts in the classification scheme to online materials at scale. There are no equivalent provisions in the Standards.

3.1. **Suggested revisions**: the guidance for classifying content should be consistent with sections 3(f), 3(g) and 5.3 of the Head Terms of the Codes to ensure a consistent approach to classification of materials between the OSA Codes and the Standards.

Specific recommendations

F. We suggest that the Standards, or at a minimum the guidance for the Standards, should provide that:

Industry participants are not required to arrange for any material to be classified or otherwise to replicate the classification functions of the classification Board. However, where an industry participant is required by a Standard to identify or distinguish between different sub-categories of unclassified class 1 material, the industry participant must develop a process for categorising that material in a way that is informed by the classification process. In each case, the process developed by the industry participant will serve as a proxy for the classification process and may vary depending on the circumstances in which it is to be applied, including to take into account the nature and volume of material to be categorised. Where the Standard is to be applied in relation to a single known item of material, the process may involve a detailed review of that material and may follow the classification process more closely than where the compliance measure is to be applied at scale in relation to multiple unknown items of material. For example, when applying a Standard at scale, an industry participant may develop particular flags or triggers in order to identify material that is likely to be class 1A or class 1B material under this Code, where those flags or triggers are framed by reference to factors that would be taken into account pursuant to the classification process. Different industry participants may develop different flags or triggers, depending on the nature of the online activities they are undertaking, the nature of the material they are dealing with, and other relevant factors.

G. We suggest that the Standards make it clear that industry participants will be treated in an equivalent way as provided for in the Head Terms i.e., that eSafety will not hold a participant in breach of a Standard because they categorise material differently to other participants or to eSafety.

4. Unclear who are 'account-holders' versus 'end-users'.

The Standards attempt to distinguish between the 'account holder' for a service, and an 'end-user' who uses that service. However, we are unclear on why this term has been introduced and this is not explained in the Discussion paper or Fact sheets. 'Account-holder' is defined in the Standards to mean in relation to a service 'the person who is the counterparty to an agreement with the provider of the service for the provision of the service.' As drafted, the definition of 'account holder' appears to capture *any end-user bound by terms of use*. Furthermore, the concept of an 'account holder' is applied to all RES and DIS equally – regardless of whether or not a particular RES or DIS service is an 'account-based' service (and many will not be, such as the majority of websites). If, for example, the object of these new terms is to distinguish between relevant categories of services based on whether the service provider allows an account-holder to add additional users to the service, this needs to be clarified. Unfortunately, we cannot make any suggestions as to how this issue should be addressed given we are uncertain what eSafety is seeking to achieve with these distinctions. We recommend that eSafety reconsider the need for this concept given the proposed approach does not appear to achieve any workable distinction and it is unclear what it is seeking to achieve.

Specific recommendation

H. We recommend that eSafety reconsider the need for this concept given the proposed approach does not appear to achieve any workable distinction and it is unclear what it is seeking to achieve.

5. Test for appropriate action: DIS Standard, section 12, RES Standard, section 12

There are a range of requirements in the Standards (as there were in the OSA Codes) for providers to take appropriate action, or otherwise justify why measures taken are 'appropriate'. However, the test of what will constitute 'appropriate' action in the OSA Codes has been replaced by a different test in the Standards.The test of appropriateness in section 5.1(b) of the Head Terms was a test of the 'reasonableness' of the provider's actions, to be assessed by balancing a range of different considerations such as privacy, and freedom of expression, the nature of the product or service in question (including its function, purpose, size/scale, maturity) and the capacity/capabilities of the provider. Significantly, the concept of 'appropriateness' in the Standards does not enable providers to take into account the range of factors that were considered critical by industry for the OSA Codes.

5.1. **Suggested revisions:** We suggest that the relevant sections of the Standards be amended so that it is clear the test of appropriateness is a test of *reasonableness in the circumstances,* taking into account the factors set out in the Standards. We also suggest that the Standards or the accompanying guidance make clear that the factors in section 5.1(b) of the Head Terms of the OSA Codes are also relevant considerations in determining whether action is appropriate 'in the circumstances'.

Specific recommendations

I. We suggest that section 12 of the Standards be amended to provide as follows:

Where the term appropriate is used to describe the actions required by the provider of a service to comply with a standard, a provider is required to demonstrate that its actions are reasonable in the circumstances, taking into account all relevant considerations, including :

a) the extent to which the action is likely *to* achieve 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 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.

J. The Standards should provide that the factors listed in section 5.1 (b) of the Head Terms of the OSA Codes are also relevant considerations in determining whether action is appropriate.

6. Detection and removal of known CSAM: DIS Standard, section 21, RES Standard, section 20

Industry remains concerned about the introduction of expansive requirements to scan for users' content on RES and DIS services. The types of services that fall within the RES and DIS categories are very diverse and providers of different service types establish trusted relationships with their users in different ways. Rather than require all services to scan for content, eSafety should provide flexibility for services to tackle the risks of CSAM and pro-terror materials that takes into account the characteristics of the services, the guarantees users rely on, and the security and privacy attributes of each service.

As foreshadowed by the eSafety Commissioner's Statement of Reasons, the draft Standards contain obligations on industry participants to implement systems, processes and technologies that detect and identify known child sexual abuse material, and remove it as soon as practicable after detection. Where this is not technically feasible,[11] a provider must take appropriate alternative action. In addition to meeting the requirements above, end-user managed hosting services and Tier 1 DIS must ensure that their services use systems, processes and technologies that automatically detect and flag known CSAM. This obligation is not limited by technical feasibility.

Question 4 of the Discussion paper asks; 'Whether the technical feasibility exception to the proactive detection obligations is appropriate, and whether there are any other limitations that would prevent service providers from being able to comply'.[12] This section of the submission explains why we think the Standards need to be amended to ensure the security of services is not compromised and to align with eSafety's stated intentions concerning the implementation of proactive detection requirements on encrypted services. In addition, we consider that the technical feasibility test should apply more broadly - for instance, the requirements to take appropriate alternative action and for Tier 1 DIS and end-user managed hosting services to use systems, processes and technologies that automatically detect and flag known CSAM should be limited by technical feasibility.

6.1. **Approach taken by industry compared to eSafety concerning proactive detection:** A key issue that arises from eSafety's proposed approach is that it extends scanning requirements to a broad range of services including private email and messaging services, and private storage services, that many Australians legitimately rely on for their privacy and security capabilities. Surveillance of many of these types of services is something that has traditionally been prohibited, restricted or protected (by way of appropriate safeguards). Industry is concerned that the Standards impose proactive detection obligations for such services in a broad fashion that does not take into account differences between services.

6.2. In addition, eSafety's proposed approach does not explicitly provide safeguards concerning the use of encryption and system security measures by RES and DIS services and only allows providers to rely on the technical feasibility test in limited circumstances. We think that the Standards should contain explicit safeguards for system security measures which are critical to safeguarding cybersecurity and users' personal information online. These safeguards were included by industry in section 6 of the Head Terms of the OSA Codes (closely based on the protections there were introduced into Part 15 of the Telecommunications Act 1997 as part of the assistance and access scheme) and should be retained in the Standards.

6.3. In our view the technical feasibility test should be applied to all relevant services that are subject to proactive detection requirements including end-user hosting services. If eSafety is not intending to require providers to do things that are not technically feasible (such as decrypting services) there shouldn't be any issue with broader application of this test.

6.4. Additionally, it is unclear how the 'technical feasibility' test will operate in practice. Arguably, it could be technically feasible for a range of services to decrypt their services in order to implement technological proactive scanning of services for known CSAM, but it might be prohibitively costly, excessively time consuming, or negatively impact a provider's overall security architecture, design, or operating system. The technical

[11] See sections 7, *RES Standard* and *DIS Standard.*

[12] *Discussion Paper*, Question 4.

feasibility test also does not clearly cover situations where providers will be unable to assess the context of materials (for example, where they do not have visibility over a chain of communications between their users and users of another service or where there are systems, processes or technology that is available, but which may be prone to errors on particular services). In addition, the draft RES Code included explicit guidance about how to deal with the risk of deploying proactive detection technologies, such as the health and safety risks to staff reviewing this material.[13] This guidance is important in ensuring that providers have adequate compliance systems for implementing proactive detection strategies that are appropriate to their business.[14]

6.5. We also note that the approach in the Standards to proactive detection is less flexible than was provided for in the OSA Codes. The SMS Code requires providers of Tier 1 social media services to *implement systems, processes or technologies that are designed to detect, flag and/or remove from the service, instances of known CSAM for example, using hashing, machine learning, artificial intelligence, or other safety technologies*. This approach was also proposed in the draft RES and DIS Code, albeit for providers of a more limited category of services than is now proposed for these Standards.[15] The rationale for a more flexible approach was to enable providers to consider a range of alternative strategies to detect and remove CSAM, other than using scanning technologies. The approach in the Standards means that RES and DIS services cannot meet the requirement to detect and remove known CSAM solely by using systems and processes, for example, by pre-moderating materials (noting that this might be the most effective approach for some services such as websites).

6.6. In addition the definition of 'known CSAM' extends to the material that has to have been verified by a recognised child protection organisation meeting relevant criteria (i.e. private organisations including NGOs) and material that has been verified by a governmental or non-governmental organisation. It would be helpful if the guidance concerning proactive detection can be clarified so that a provider can choose one or more verified lists – rather than having to ensure it covers all possible governmental and NGO lists.

6.7. **Suggested revisions:** Industry considers the proactive detection requirements should not extend beyond what was proposed in the draft RES and DIS Codes. To the extent that these provisions are retained, we think the technical feasibility test should be applied to the proactive detection obligations broadly (including requirements to take appropriate alternative action), including for end-user managed hosting services. We think the technical feasibility test should be applied to the proactive detection obligations on end-user hosting services. We understand that eSafety does not intend to require service providers to monitor the content of private emails, instant messages, SMS, MMS, online chats and other private communications. Nor does eSafety 'expect companies to design systemic vulnerabilities or weaknesses into end-to-end-encrypted services'.[16] We think

[13] See guidance for section 8, *SMS Code*.

[14] See concept of "capable of reviewing and assessing material" as used in the draft RES and DIS Codes. This concept covered not just technical capability, but also a provider's legal and practical ability to review and assess material on a service. See further comments in part 17 of this submission.

[15] Under the draft *RES Code* this was required of a provider of: a) a Tier 1 relevant electronic service; b) an open communication relevant electronic service that is not a carriage service provider; c) a dating service; or d) a gaming service with communications functionality. Under the DIS Code, this was required of a Tier 1 Designated Internet Service.

[16] *Fact Sheet, Draft Online Safety (Relevant Electronic Services – class 1A and class 1B Material) Industry Standard 2024,* Office of the eSafety Commissioner, November 2023.

that these statements need to be clearly and explicitly incorporated in substantive provisions of both the RES and the DIS Standards, consistent with section 6.1 of the Head Terms for the OSA Codes. The 'technical feasibility' test should be revised to be much clearer that all relevant considerations going to feasibility can be considered. In addition, the guidance for implementing proactive detection measures should closely align with that in the Social Media Services Code (SMS Code). We would also suggest consideration be given to amending the proactive detection requirement to allow industry participants the flexibility to implement systems, processes or technologies that are designed to detect, flag and/or remove from the service, instances of known CSAM to ensure a consistent approach to this measure across RES, DIS, and SMS services.

Specific recommendations

K. The Standards should provide flexibility for providers of services to best tackle the risks of CSAM and pro-terror materials in a manner that is appropriate for individual services given the broad range of service types covered by these Standards, and the complex privacy and security considerations that accompany many of these services.

To the extent that provisions concerning proactive detection obligations for CSAM the following Recommendations are also made:

L. The Standards should be more broadly subject to the technical feasibility test. For instance, all detection obligations, obligations contingent on visibility of content/communications and other obligations requiring technological solutions in the Standards (including requirements to take appropriate alternative action) should be subject to the technical feasibility test. For example, the proactive detection provisions in the DIS Standard relating to end-user hosting services should be subject to the technical feasibility test and providers should not be required to take "appropriate alternative action" that is not technically feasible.

M. The proactive detection requirement should allow industry participants the flexibility to implement systems, processes or technologies that are designed to detect, flag and/or remove from the service, instances of known CSAM to ensure a consistent approach to this measure across RES, DIS, and SMS services.

N. The Standards should be amended to make clear that they do not require any industry participant to undertake steps that do the following (regardless of the costs involved):

a) implement or build a systematic weakness, or a systematic vulnerability, into a form of encrypted service or other information security measure such as private firewall configurations, VPN tunnels and private networking links, which work directly or complement encryption to protect legitimate cybersecurity and data integrity interests;

b) build a new decryption capability in relation to encrypted services;

c) render methods of encryption less effective; or

d) undertake monitoring of private communications between end-users.

O. The considerations that should be taken into account in applying the technical feasibility test in sections 7 of the Standards should include a new considerations: '(c) whether it is reasonably practical for the provider to review or remove material on the service' and '(d) the accuracy and reliability of the available proactive detection systems, processed or technologies'. In addition, the Standards should be revised to clarify that cost is only one consideration for technical feasibility (i.e., lower cost is not viewed as a controlling factor, by itself, that an action is feasible). The Standards should reflect the fact that technical feasibility encompasses a range of concerns going to feasibility – not just whether something is technically possible or feasible, but also whether it is (i) reasonable taking into account concerns such as security and privacy risks (eg whether it would negatively impact a provider's security architecture, design or operating system); and (ii) feasible from a legal and practical perspective (see comments on part 17 of this submission below).

P. We suggest that the additional requirements for end-user managed hosting services and Tier 1 (see section 21(8)) are unnecessary given the remainder of the measure.

Q. Clarity should be included regarding what 'appropriate alternative action' could be for fully end-to-end encrypted services.

R. The guidance for the proactive detection provisions should direct providers to consider:

   a) the availability of different options for proactive-detection of known CSAM and the capability of the provider to use those options accurately and reliably;
   b) the need for human resourcing required to review detected materials, for example, where an image that has been detected has been altered to make it more difficult for it to be matched with the original image;
   c) the need to provide adequate health and safety arrangements for personnel undertaking the review of known CSAM; and
   d) the fact that hash lists of known CSAM are not infallible. Providers should therefore take care to safeguard against low quality hashes and hashes prone to collisions (e.g., compilation videos) by having a suitable confirmation and

quality control process to independently confirm that the material depicted in the hash is CSAM. Where a hash is likely to lead to false results, a provider should not deploy it.

7. Detecting and removing known pro-terror material: DIS Standard, section 22, RES Standard, section 21

The draft Standards contain obligations on industry participants to implement systems, processes and technologies that detect and identify known pro-terror material, and remove it as soon as practicable after detection. Where this is not technically feasible, the provider must take appropriate alternative action.

The approach in the Standards to proactive detection of known pro-terror materials raises the same issues identified for proactive detection of known CSAM, discussed in section 6 of this submission. In addition it appears that industry and eSafety have different understandings of what types of material can be treated as known pro-terror material, and the extent that material can be accurately detected online.

The approach of the Standards appears to assume that pro-terror materials can be accurately detected by providers of RES and DIS services using reliable, specialist databases of pro-terror material, without deploying a trained human reviewer to analyse the context of the relevant material. We consider that this is inconsistent with the OSA, which requires online material to be treated in an equivalent manner to offline materials under the classification Scheme.The classification scheme provides that material is only pro-terror material if it:

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

b) *it directly or indirectly provides instruction on the doing of a terrorist act; or*

c) *it directly praises the doing of a terrorist act in circumstances where there is a substantial risk that such praise might have the effect of leading a person (regardless of his or her age or any mental impairment (within the meaning of section 7.3 of the Criminal Code) that the person might suffer) to engage in a terrorist act.*[17]

We interpret this to mean that material is only pro-terror material if it is encouraging an identifiable person, class of persons or the public to engage in terrorist acts. In other words, the approach of the classification scheme is predicated on the fact that the harm from pro-terror material arises from its dissemination and consumption, and not per se in the creation of material that could potentially be harmful (or not). This assessment of whether material is pro-terror necessarily requires an analysis of the context of the material and, even then, the assessment may be highly contestable amongst different international bodies and governments. For example, the Australian government has taken a different position to the UN in respect of certain material that has been distributed online in the current conflict in Gaza and Israel. Furthermore, the classification scheme acknowledges that there are a range of contextual factors that need to be considered in assessing materials. For example, material does not advocate the doing of a terrorist act if it ‘depicts or describes a terrorist act, but the depiction or description could

[17] Sec 9A (3) of the *Classification Act.*

reasonably be considered to be done merely as part of public discussion or debate or as entertainment or satire'.[18]

The need for DIS and RES providers to assess the context of material before it can be assessed as pro-terror material within the meaning of the OSA creates a range of practical issues for industry participants implementing the Standards for proactive detection of pro-terror materials.

7.1. **Challenges for proactive detection of known pro-terror materials on DIS and RES services:** Question 6 of the Discussion paper asks: are there any limitations which would prevent certain service providers from deploying systems, processes and technologies to disrupt and deter child sexual abuse material and pro-terror material on relevant electronic services? In industry's view, proactive detection of pro-terror materials on RES and DIS services cannot be carried out in the same way as proactive detection of known CSAM. It is possible to engage in proactive detection of known CSAM at scale because CSAM is always harmful in any context and there are highly reliable databases established for that purpose by internationally recognised non-governmental bodies such as the National Centre for Missing and Exploited Children. The nature of known CSAM is that once detected it is usually readily categorised as such.[19] In contrast, there is no equivalent independent expert authority anywhere in the world that maintains a database of verified pro-terror material (as defined by the OSA) that can be reliably used to proactively detect 'known pro-terror material' online. Again, we note the crucial importance of context in verifying material as pro-terror material. There are, however, a few organisations such as Tech Against Terrorism and the Global Internet Forum To Counter Terrorism that provide indicators of material which may, in a given context, be assessed by providers of online services to be pro-terror materials. However, these are independent organisations with strict membership criteria and processes that are intended to ensure users of their databases are equipped with appropriate systems and processes for the review of detected materials, to ensure fundamental human rights are upheld. It follows that, to comply with the requirements in the Standards for proactive detection of pro-terror materials, providers will have a much greater need for trained human reviewers to undertake analysis of the context of the material to determine if it is, in fact, pro-terror material.

7.2. In the case of 'inert' file storage on end-user hosting services, or in email 'drafts' folders, it is practically impossible for service providers to make an accurate assessment of stored material, as the purpose for which material is stored will be unknown by the provider of the service. It is only when material is shared with an audience that is likely to cause harm, or it can potentially be assessed by a provider as encouraging/instructing persons to take action of the kind described in the definition of pro-terror material. For example, a manifesto may be stored by an end-user for educational purposes, scholarly research or journalism. Such material can only be accurately assessed as pro-terror material when it is sent to or shared with another person as part of a call to promote a terrorist act.

7.3. **Suggested revisions:** In addition to the recommendations in section 6 of the submission above being extended to pro-terror material, we suggest that the obligation to detect pro-terror material should only apply to DIS or RES services to the extent that material is

[18] This exemption in the classification Act which requires an analysis of the 'depiction or descriptions of the material'. This formula is not exactly replicated in the definitions of pro-terror material in the Standards, which require an analysis of 'the availability on the service' of the materials.

[19] As noted above, there may still be circumstances where human verification of known CSAM is required , for example, where an image has been altered to make it more difficult for it to be matched with the original image.

shared or sent to another person and not to materials that are stored on the service in an inert state. Further the effective deployment of proactive detection technology to identify and remove pro-terror material requires the deployment of trained human reviewers. This may impose an unreasonable burden on small to medium size services. We suggest that proactive-detection of pro-terror materials only be required of those services that have more than 2 million active Australian end-users.

Specific recommendations

Please see Recommendation K, which also applies to pro-terror materials. To the extent that provisions extending proactive detection obligations for proactive detection of pro-terror material, the following Recommendations are also made:

S. Recommendations L, M, N, O, Q and R should also apply to pro-terror materials.

T. For consistency with the definitions of pro-terror material in the classification scheme, the Standards should be amended to make clear that the requirements to proactively detect pro-terror materials apply only to the extent that the material has been sent to or shared with another person via the relevant DIS or RES service and not to material that is stored on the service in an ‘inert state’.

U. The Standards should limit the requirement to proactively detect known pro-terror material to relevant RES and DIS services that have more than 2 million active Australian end-users so as to not impose an unreasonable cost burden on small to medium size businesses who would need to hire specialised moderators to comply with the requirements.

V. The Standards should be accompanied by guidance that makes clear that providers need to ensure detected material is subject to human review before it is classified a pro-terror material in accordance with the requirements of the classification scheme. In addition similar guidance needs to be provided as for known CSAM concerning the compliance systems that need to be in place e.g to prioritise material for human review and protect the health and safety of personnel that are reviewing potential pro-terror material.

W. The absence of any organisation that can verify whether material is pro-terror material or the definition of pro-terror material is a significant issue for industry. As drafted, it is not clear how eSafety expects industry to access ‘verified material’,, which is of real concern as different nation states and international bodies such as the UN take widely different approaches to determining what material is terrorist in nature. We suggest the Standards should be amended to make clear that ‘known-pro-terror material’ includes material that eSafety has previously verified as pro-terror material consistent with the approach of the OSA Codes. Further consideration should be given to establishing systems and processes within eSafety for verifying pro-terror material including in cooperation with the Classification Board, and making its decisions about pro-terror material available to industry.

8. Notification of CSAM and pro-terror materials to ‘verifying organisations’: DIS Standard, section 15 and RES Standard, section 15

The RES and DIS Standards contain additional requirements on industry participants to notify certain ‘organisations’ that can verify if material is pro-terror material or CSAM in circumstances where it has identified material on the service that it believes in good faith is ‘not known CSAM’ and ‘not known pro-terror material.’[20] The Standards do not clearly define the types of authority which are the subject of the notification requirements. It isn't clear what the notification measures for RES and DIS aim to achieve and why they are significantly higher/more onerous in a range of regards than other categories of service regulated by the OSA/Codes.

8.1. **Challenges of notifying CSAM and pro-terror material to verifying organisations:** We are unaware of any organisation which can *verify* material as CSAM or pro-terror material under the classification scheme, other than eSafety or the National Classification Board (including on referral by eSafety under the OSA)[21] Furthermore, in Australia there is no central authority that has legislated responsibility for receiving reports of materials and maintaining an updated repository of verified pro-terror materials.

8.2. It appears that the Standards do not intend that industry participants notify eSafety or the Classification Board of first generation pro-terror and CSAM, but expect industry to notify NGOs outside of Australia, which are governed by foreign laws. In the case of CSAM, industry participants that are regulated under US law, are obliged to report CSAM material to the National Center for Missing and Exploited Children, including potential new generation material on the service. One of the key aims of the scheme is that it strengthens NCMEC’s databases of hashed material that can be utilised by law enforcement and certain approved industry participants to proactively detect CSAM materials. There are also additional databases of CSAM materials operated by other expert authorities such as the Internet Watch Foundation (IWF) which may be accessible by industry participants. However, the difficulty is that these organisations cannot verify materials in accordance with the OSA. It is also unclear whether NGOs such as NCMEC or IWF, are obliged to make these databases available to foreign based companies regulated by the OSA. In the case of pro-terror material, we are unaware of any equivalent non governmental organisations to NCMEC that maintains a database of verified pro-terror material under the classification scheme or otherwise.

[20] Sec 15(3) and (4) of RES Standard and section 12(3) and (4) of DIS Standard.

[21] We understand that rather than ‘verifying material’ in accordance with the laws of the different jurisdictions from which it receives reports, child protection organisations, such as NCMEC evaluate reports they receive, in accordance with its own processes and may, in their discretion, decide to refer the case to a relevant law enforcement agency in Australia or elsewhere.The local law enforcement agency has discretion to choose to investigate NGOs reports further.

Specific recommendations

X. We think that the obligations to notify a verifying authority for CSAM and pro-terror materials should be removed given there are no authorities that maintain databases of verified CSAM or pro-terror materials under Australian law. Alternatively the obligation should be limited to notifying materials that a provider reasonably believes are not known CSAM to a non- governmental expert child protection authority which maintains a database of materials that may be accessed by industry participants subject to the Standards in order to proactively detect CSAM on their services.

9. Disrupting and deterring CSAM and pro-terror material: DIS Standard, section 23

The Standards contain new requirements for DIS and RES providers to disrupt and deter pro-terror materials, which substantially differ from the equivalent disrupt and deter obligations in the SMS Code.[22] It is unclear why the RES and DIS Standards warrant such a significantly higher and inflexible bar of compliance compared to the other service categories under the Codes and the OSA. The key differences between the approach in the Standards and the equivalent measure in the SMS Code are that:

a) the SMS Code requires that industry participants invest in systems, processes and/or technologies that aim to disrupt and deter pro-terror material – all three actions are required under the Standards;

b) the new requirements for DIS and RES providers extend to enterprise services; whereas enterprise services were exempted from this requirement in the draft OSA Codes. Enterprise providers generally do not have the technical or legal capabilities to disrupt discrete instances of content, but rather only possess blunt instruments such as suspending the customer's entire service. Additionally, for some enterprise customers such as government, law enforcement, journalists, legal professionals and human rights organisations, possession of certain materials may be required.

c) the new requirements are less flexible than the SMS Code. The examples of 'disrupt and deter' compliance measures in the SMS Code make clear that compliance can be achieved in a variety of ways, including by funding research into new technological solutions; and

d) the equivalent measure in the SMS Code is designed to ensure that industry commits to continuously improving its efforts to combat CSAM and pro-terror material, particularly novel or 'first generation' material, in a way that is proportionate to the risk of that material surfacing on a service. However, the DIS and RES require providers to implement systems, processes and technologies to *effectively* disrupt and deter end-users from using the service to disseminate CSAM/pro-terror material. Importantly, the Standard is

[22] See MCM 10, SMS Code.

not limited by technical feasibility or reasonableness or appropriateness in the same way as the two detection and removal obligations in the Standards for known pro-terror material and known CSAM, as set out above. As drafted, the Standards will likely result in services implementing AI based technologies (being the most effective available technologies) to meet the disrupt and deter requirements for CSAM/pro-terror material even though many of these technologies are at an early stage of development and cannot be reliably deployed without the support of trained reviewers.

In addition, we note that the draft DIS Code did not impose obligations on end-user hosted services to deploy technology to identify and remove pro-terror material because it is not practically possible to identify pro-terror material (whether 'known' or 'unknown') on inert file storage, as discussed above.

9.1. **Challenges of disrupting and deterring pro-terror and CSAM material:** As drafted, the new disrupt and deter approach to CSAM and pro-terror materials for DIS and RES services set an unachievable bar for compliance for many businesses in scope of the Standards. We are unaware of any systems, processes and technology which will be 100% effective in detecting this material online. We think consideration needs to be given to setting a more realistic and flexible standard of compliance and limiting the scope of these provisions in a way that is proportionate to the risk and capabilities of different types of services.

9.2. Question 20 of the Discussion paper asks in relation to relevant enterprise providers, do the proposed obligations (in particular, the section 23 obligation to 'disrupt and deter') provide appropriate safeguards? Are there specific challenges to deploying these measures? We think that the expansion of requirements to disrupt and deter pro-terror and CSAM are likely to be problematic for many enterprise providers who will be contractually bound not to interfere with the information and data stored by customers on communications and file storage services.

9.3. While the accuracy of technology to enable detection of first-generation CSAM and potential pro-terror material is improving, it is generally accepted that it is not as accurate as technology for the detection of known CSAM. Even with trained human reviewers, it can be difficult for providers of very large DIS and RES services to accurately assess the context of material on many DIS and RES services.[23] We think that the 'effectiveness' bar for compliance is especially inappropriate for encrypted services, email services and small to medium sized businesses given that the most effective AI based technology is at an early stage of development, cannot be deployed without review of detected materials by human assessors, and will be cost prohibitive and /or not technically feasible for some providers to implement. As discussed in section 7 of the submission above, in the case of end-user hosted services and stored material on email services, it is practically impossible for providers to assess whether material is pro-terror when it is in an inert state. Many enterprise DIS services, such as cloud storage services provided to business and government, may also be unable to comply with the Standard where the terms of their customer agreements prohibit the deployment of technologies of the kind contemplated by the Standards.

9.4. **Suggested revisions:** We suggest that revisions be made to the Standards and accompanying guidance to ensure that the requirements on industry to disrupt and deter

[23] See https://www.nytimes.com/2023/11/27/technology/google-youtube-abuse-mistake.html

CSAM and pro-terror material enables providers the flexibility to implement a range of systems, processes *or* technologies' in a manner that is commensurate with the risk of those materials surfacing on the service and 'where technically feasible'.[24] We also consider that the requirement should be to deploy 'appropriate', rather than 'effective' systems, given that none of the currently available solutions are 100% effective. The disrupt and deter requirements should not apply to services with less than 2 million active end-users, or to enterprise services at all, and no RES and DIS services should be required to disrupt and deter pro-terror material to the extent it is in an 'inert state'.

Specific recommendations

Please see Recommendation K, which also applies to the obligations to disrupt or deter CSAM and pro-terror material. To the extent that the Standard includes disrupt and deter provisions, the following Recommendations are also made:

Y. We suggest that these disrupt and deter requirements in the Standards for CSAM and pro-terror material be amended so that:

a) The measure: (i) provides more flexibility for providers to achieve compliance in a number of ways – given the range of providers covered and the early stage of relevant technologies; and (ii) ensures that industry commits to continuously improving efforts in a way that is proportionate to the risk without requiring the implementation of effective tools as at day one (which is an unrealistic bar, given that many of these technologies are still early stage for a range of service offerings and not yet effective).

b) The requirement to implement 'effective systems, processes and technologies that disrupt and deter CSAM and pro-terror material' is replaced by a requirement by reference to 'appropriate systems, processes and/or technologies that disrupt and deter CSAM and pro-terror material'.[25]

c) If the requirement to implement technology is retained, it should be limited to providers with more than 2 million active end-users in circumstances where there is a reasonable suspicion there is new generation (unknown) CSAM or pro-terror material on the service.

d) The technical feasibility test should apply to this measure, including as modified by L, N and O above (i.e., providers are not required to take any action that is not technically feasible);

e) Enterprise services should be excluded from the disrupt and deter pro-terror requirements; and

f) The requirements to disrupt and deter pro-terror material should not apply to the relevant DIS or RES services to the extent that material is stored on the service in an inert state, and is not accessible to another person.

[24] See discussion of technical feasibility at section 6 of the submission above.

[25] See discussion on need to revise the meaning of 'appropriate' in section 5 of the submission above.

Z. eSafety should also provide detailed guidance that explains the risk of 'false positives" and how these risks should be mitigated by providers, including guidance concerning the need for the provider to make suitable arrangements for the human review of detected materials and the need to accurately verify the context of the detected materials before enforcement action is taken under the providers terms of use.

10. Development programs, DIS Standard, section 24, 34(2)(d) and 37, RES Standard, section 23, 33(1)(d) and 36

The Standards require certain providers meeting minimum user numbers to establish and implement, each calendar year, 'a program of investment and development activities for systems, processes and technologies' meeting certain minimum requirements. These obligations for RES apply to Tier 1 and Pre-assessed RES whose average monthly number of active end-users of the service in Australia is 1,000,000 or more. For DIS, the obligations extend to Tier 1 DIS and Generative AI DIS 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 to end-user managed hosting services 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. Such providers can also be required to provide their development programs for a calendar year to eSafety, and to subsequently report to eSafety on the outcomes of the development program for a calendar year.

10.1. **Industry concerns about new requirements for development programs:** We question whether this provision is appropriate or proportionate. In our view, eSafety should be primarily focused on whether the provider is achieving the requirements of the Standards at any given point in time, not the provider's future plans. It is unclear to us why such an onerous new and additional requirement is needed for RES and DIS over and above the requirements for the other categories of service regulated by the OSA Codes. Furthermore section 37 and 36 are unnecessarily duplicative of the requirements imposed on providers under section 56 and 58 of the OSA, which already provide mechanisms for eSafety to obtain relevant information about providers' development actions in relation to class 1 materials, both on an ad hoc and periodic basis.

10.2. **Suggested revisions:** We suggest that the requirements for development programs be removed. At a minimum, if retained they should be amended to ensure that they are sufficiently flexible to enable providers to develop systems processes *or* technologies that are proportionate to the platforms' assessment of the risk of class 1A or class 1B materials on their platform. In addition we suggest that the requirements of this obligation should only apply to services with more than 2 million active monthly end-users.

Specific recommendations

AA. We suggest that the requirements for providers to establish a development program should be removed.

BB. If retained, at a minimum, they should be amended as follows:

a) section 24 (3) of the DIS Standard and section 23 (3) of the RES Standard should be amended to provide that 'the provider of the service must establish and implement, for the calendar year, a program of development (***development program***) in respect of systems, processes or technologies that is *reasonably proportionate to the risk of class 1 material on the service* as assessed in accordance with this Standard;

b) the mandatory requirements should be discretionary i.e., the word 'must' in section 23(4) of RES Standard and section 24(4) of the DIS standard should be replaced by 'may'; and

c) The requirements of this obligation should only apply to services with more than 2 million active monthly end-users.

CC. Section 37 of the DIS Standard and 36 of the RES Standard should be deleted.

11. Terms of use, DIS Standard, section 14, and RES Standard, section 14

Section 14 of the Standards sets out new requirements for:

a) provisions to be included in the terms of use for most DIS and RES;

b) obligations to enforce those terms of use when it becomes aware of a 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; and

c) provides that the provider will bear an evidential burden of proof in proceedings to establish that it did enforce its terms, and that it did so in a way that was appropriate and proportionate.

Whilst the OSA Codes include provisions requiring providers to have terms of use that address class 1 materials, these are markedly different to the requirements for terms of use in the Standards. It is unclear to us why eSafty considers it necessary and proportionate to take a more onerous approach to RES and DIS services in comparison to other service categories under the OSA Codes.

In addition, we think that it is important that providers have a degree of discretion to take appropriate forms of enforcement action for materials that are not per-se illegal. Appropriate enforcement action could include a broad range of other steps in addition to the options mentioned in 14(2)(b).

11.1. **Challenges of implementing requirements for terms and conditions:** As drafted the Standards in effect, require all class 1A and class 1B material to be prohibited by regulated service providers. In contrast, the Codes enable providers to take a different approach to class 1A and 1B material because class 1B content and, also some extreme crime and violence material falling within class 1A, is not per se unlawful offline in all circumstances.

11.2. Under the classification scheme it is clear that these materials must be evaluated in relation to the context in which the material appears. For example, drug-related material may be permissible where used in an anti-drug context.

11.3. Further, some dealings with some forms of content such as drug related and crime and violence material are lawful offline (e.g., merely holding a copy of material). These distinctions were sought to be maintained in the Codes. For example, the draft DIS Code acknowledged (for end-user managed hosting services) that some class 1B and class 1A crime and violence material could be permissible on a storage service in certain circumstances.

11.4. There seems to be inconsistency across and between the two Standards as to which sub-categories have to be prohibited and removed. Section 14 also seems inconsistent with other provisions in the Standards (e.g., section 32 of the DIS Standard) which *does* maintain the distinction between the lawful and unlawful material categories . It is unclear why the approach to class 1B in section 25 of the RES Code requires removal of class 1B materials whereas the equivalent provision of the DIS Standard does not (it simply requires appropriate action).

11.5. **Suggested revisions:** In our view, it is important that class 1 materials are broadly treated the same way online and offline. We further recommend that the Standards should deal with the different sub-categories of class 1 materials in a consistent manner: CSEM and pro-terror materials should be prohibited as these are per-se unlawful, other categories of class 1 materials should be subject to a more flexible approach that requires providers to have 'appropriate terms and conditions'. The inconsistency in treatment of the same subcategories of materials (e.g., class 1B materials) between DIS and RES should also be rectified.

11.6. The obligations on providers to ensure terms and conditions contain specific rights to take enforcement action in section 14(3) are unduly prescriptive, particularly with respect to class 1 materials which are not per se unlawful. We think section 14(3)(b) should be limited to CSAM and pro-terror materials. In addition, we think that it is important to acknowledge that the ability of a provider to identify a breach turns on technical feasibility (and also legal and practical feasibility) as highlighted in section 6 of the submission above. We think that section 14(3) should be amended to acknowledge this and to make clear that the action taken should be 'reasonably proportionate' to the breach.

11.7. There also appears to be a requirement to oblige account holders to ensure that end-users in Australia broadly (not just end-users that the account-holder authorised to use their account) comply with terms and conditions. We assume that this is a drafting error. As noted in section 4 of this submission above, it isn't clear what the concept of an account-holder is intended to achieve or, for example, how an account-holder differs from an end-user. See recommendation H.

11.8. In addition, section 14 imposes a requirement on enterprise services to obtain rights under their customer agreements to suspend the provision of the service to a specified end-user of the service for a specified period if they breach the prohibition on class 1 materials. This is not practically possible under the terms of many enterprise agreements and should be removed.

11.9. Lastly, we are unclear why the evidentiary burden for compliance under the Standards should be different to the Codes or to the enforcement powers given to the Commissioner under the OSA, which both require eSafety to make out a breach. We therefore recommend that the change to the evidentiary burden be removed.

Specific recommendations

DD. We suggest that the requirements in sections 14 be amended as set out below:

a) The primary requirement in sections 14(2) should provide that ‘The provider of a service must include provisions in the terms of use for the service that:

i) prohibit end-users from using the service for CSEM or pro-terror material; and

ii) contain appropriate restrictions on end-users concerning other categories of class 1 materials’.

b) The inconsistencies in the treatment of class 1A material other than CSEM and pro-terror material, and of class 1B material, between the RES and DIS Standards should be rectified.

c) The requirement in section 14(2)(b) should be limited to CSAM and pro-terror materials.

d) The requirement on enterprise services in section 14(2)(b)(ii) to include provisions that enable them to suspend the provision of the service to a specified end-user of the service for a specified period if they breach the prohibition on class 1 materials should be removed.

e) Section 14(3) should be amended as follows:

> Where technically feasible, where 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 reasonably proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach.

The changes to the technical feasibility test recommended in part 6 of this submission above should also apply here. Please also see recommendations in part 17 of this submission below.

f) The changes to the evidentiary burden in sections 14(4) should be removed.

g) That the additional requirement for end-user managed hosting services in section 16(6) of the DIS Standard is either removed (as it is unnecessary) or that clarity is provided as to what else is required by section 16(6) in addition to what is already covered by section 14.

12. Transition period: DIS Standard, section 2, RES Standard, section 2

The OSA Codes granted industry a 6 month transition period to implement the required measures plus a second 6 month period during which participants have leeway to implement specific measures where there were reasonable grounds for not being fully compliant (e.g., significant engineering or system changes) provided the participant can demonstrate to eSafety's reasonable satisfaction that it is working towards achieving compliance by the end of that second 6 month period.[26] There is no provision for that second 6 month period in the Standards.

12.1. **Challenges of implementing requirements for terms and conditions:** The 6 month transition period (including for any significant engineering or system changes) will be very tight for many businesses, particularly given that the Standards are markedly different from the draft Codes for DIS and RES. It is likely that implementation of the Standards will require many participants to make changes to their current systems and processes, including engineering changes that, realistically, will take 12 months.

12.2. **Suggested revisions:** We suggest that eSafety amend the Standards to make allowance for an additional 6 month period for services to be fully compliant, consistent with section 7.2(b) (iii) of the Head Terms of the OSA Codes.

Specific recommendations

EE. The Standards be amended to make allowance for an additional 6 month period for services to be fully compliant, consistent with section 7.2(b) (iii) of the Head Terms of the OSA Codes.

13. Treatment of enterprise customers: DIS Standard, definition of 'enterprise customer' and section 14, RES Standard, definition of 'enterprise customer' and section 14

The compliance measures in the draft OSA Codes applied solely to the provider of the enterprise DIS/RES. A note has been added into the DIS Standard under the definition of 'enterprise customer' stating 'The enterprise customer will often make the service available to a class of end-users, such as its staff'. When the note is read with the new definition of 'provide' this

[26] See section 7.1(b) of the *Head Terms* which provides ' If after the effective date in section 7.1(a) 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'.

suggests that enterprise customers may themselves be regulated under the DIS Code as service providers. We are unclear what eSafety's intention is, but this needs clarification noting that the OSA enables Standards to be imposed on providers of services only.

14. Prohibition on services increasing risk of class 1 materials: section 8(5) of RES standard and section 8(5) of DIS standard

The Standards contain new prohibitions on providers of a RES or DIS services making material changes to services unless:

a) a risk assessment of the service has been carried out; or

b) the change will not 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.

These requirements assume that any dealings with class 1 materials online are by definition harmful and should be prohibited online. As discussed in section 11 of the submission above, not all categories of class 1A and 1B material are unlawful (online or offline): only CSEM and pro-terror material. Certain dealings with most types of class 1 materials are permissible. For example, it may be legally permissible to store and transmit class 1A and class 1B materials (other than CSEM) for certain purposes such as research and journalism. In relation to other categories of class 1A and class 1B materials, the restriction on changes to the service should only apply to the extent that a change is *reasonably likely to materially increase the risk that materials will cause harm to other end- users of the service in Australia*.

Specific recommendations

FF. We suggest that prohibition on changes to a service that increase risk for class 1A and IB materials (other than CSEM) applies only where the change materially increases the risk that such materials will cause harm to other end-users of the service in Australia.The Standards should also make clear that employing new or different encryption or other information security measures shall not be deemed to increase the risk for class 1A and 1B materials.

15. Risk assessments for ‘deemed categories of services’: DIS Standard, section 8(5), (6) and (7) and section 25, RES Standard, section 8(5), (6) and (8) and section 19

It appears the requirements to undertake risk assessments for changes to a service now apply to all designated internet services and relevant electronic services, including those that are in ‘deemed categories’ where the applicable requirements are specifically assigned to the service. Further, it appears that following completion of any risk assessment, the provider of a deemed category of service is required to assign a risk profile under section 8(7) of the DIS Standard and section 8(8) of the RES Standard.

Under the draft DIS and RES Codes, the requirement to undertake a risk assessment was only triggered if the change to the service could either result in the service falling within a higher risk tier, or in the case of a deemed category like an end-user managed hosting service or a closed communications RES, take it outside that category.

15.1. **Issues with approach to risk assessment for 'material changes' for pre-assessed service categories.** Question 1 of the Discussion paper asks: Are the requirements for risk assessment in the draft Standards targeted at the right services and at the right points in a service's development journey? Are the risk factors appropriate? We are of the view that the approach for risk assessments upon changes to the service is inappropriate and confusing where the services are in 'deemed categories' because the compliance requirements set by the standards are assigned to those categories regardless of the results of the risk assessment. For instance, if a provider has to do a risk assessment due to material change in the service, does it have to assign a risk profile even though it still falls into the same category of service? If so, which category would apply?

15.2. In our view, if a risk assessment makes no difference to the application of the Standards to the service, this requirement should be redundant. The requirement to undertake a risk assessment should apply to a deemed category service only where it makes a material change that results in the service becoming a different category of service. For example, an end-user managed hosting service that made a material change to the service that carried a minor increase in risk would have to do a risk assessment even though, it would still be an end-user managed hosting service afterwards (and, if being launched as a new service with that change, it would not have to conduct a risk assessment for launch because of section 8(6)). The same example can be applied to closed communication RES under the RES Standard.

Specific recommendations

GG. We suggest that the Standards be amended so that services that have pre-allocated compliance requirements under the Standards are only required to carry out risk assessments when they make a material change in the service that changes the way the service is categorised under the Standard (and thus the requirements that apply to the service under the Standard).

16. Community standards: DIS Standard (see, for example, section 17), RES Standard (see, for example, section 17)

In addition to responding to breaches of terms of use and acceptable use policies, many requirements in the Standards also require providers to apply and enforce 'community standards'. There is also an additional requirement to document community standards in section 32(2)(b) of the DIS Standard and 31(2)(b) of the RES Standard.

16.1. **Issues with requirements concerning community standards:** The term 'community standards' is undefined and it is unclear whether these are limited to community standards formulated by the provider (which, in the OSA codes are treated as equivalent to terms and conditions) or some broader conception of 'community standards' under the National Classification Scheme.

Specific recommendations

HH. The term 'community standards' should be removed to avoid confusion if it is intended to be equivalent to terms and conditions and acceptable use policies (as per the OSA Codes). The term 'community standards' should be replaced by another clearly defined term in the Standards if this term is intended to cover something different to terms and conditions and acceptable use policies.

17. Engagement with end-user content reports: DIS Standard (various provisions) and RES Standard (various provisions)

Most providers, including providers of encrypted services, subject to the Standards are obliged to review/engage with content reports from end-users and take specified action to assess and respond to breaches of terms of use as a result (including by taking required enforcement action). This departs from the approach suggested in the draft DIS and RES Codes that acknowledged that different providers would have different capabilities to assess and review materials because of legislative, contractual and technical constraints. Both the draft DIS and RES Codes structured a range of measures around the concept of a provider being 'capable of reviewing and assessing or removing material' or not. Some of the measures (those involving content review and assessment against terms of use etc) were drafted with differing obligations based on that capacity. It meant that these types of provisions were structured so that providers that did have that capability would undertake that review and act accordingly pursuant to minimum measures, whereas those that did not have that capacity had much more limited obligations (e.g., to make referrals to eSafety). This concept has not been retained in the Standards.

17.1. **Issues with requirements concerning user reports:** Question 9 of the Discussion paper asks: Are the end-user reporting requirements workable for the relevant service providers? Are there practical barriers to implementation? In our view, as drafted the requirements are difficult to interpret and raise difficult questions for providers (including for encrypted services).

17.2. Firstly, it is unclear in what circumstances a provider will become 'aware' that there has been a breach of terms of use, acceptable use policies or community standards of their service. For example, it is unclear whether a provider will be 'aware' of a breach in circumstances where an end-user merely makes report of material they claim to have seen on the service, or if an end-user provides screenshots of material even if, in either case the provider has no ability to investigate and independently verify if that report is valid (e.g., due to encryption or because the provider is unable to access information needed to properly assess the context of the material).

17.3. Secondly, as drafted, some of the obligations to respond and take action in relation to reports are subject to a technical feasibility requirement, but others are not.

17.4. Thirdly, the removal requirements are different for DIS and RES providers. The reasons for these differences are not explained in the materials accompanying the Standards. For example, under section 17 of the RES Code, relevant services (including closed communication RES) who become aware of a breach of terms of use/acceptable use policies/community standards in relation to class 1A material must remove the material, unless this is not technically feasible. The provider must also take appropriate action to ensure the service no longer permits access to or distribution of the material at all. In contrast, under section 17 of the DIS Code, relevant services (including end-user managed hosting services) who become aware of a breach in relation to CSEM and pro-terror material must remove the material (this is not limited by technical feasibility in the same way as RES) and terminate user accounts in certain prescribed circumstances, as well as taking appropriate action to ensure that the service no longer permits access to or distribution of the material at all.

17.5. This approach implies that all providers must take all necessary action (including actions that will compromise encryption on their services) in order to remove certain types of material and take other types of enforcement action, even if, in practice, they cannot verify whether a report is validly made.

17.6. As drafted the requirements to remove material arguably extend to all other instances of material on the service (not only materials that the provider is aware of) and sets an extremely high bar for content removal which is unlikely to be feasible for many services to achieve.

Specific recommendations

II. We suggest that the Standards be amended so that it is clear the requirements to respond to reports and take specific enforcement action, including removal of material, only apply where the provider is aware that the specific instances of the material is in breach of its terms and condition and, in relation to materials that are the subject of a report, it is technically feasible for the provider to investigate and independently verify the accuracy of the report. See also: discussion and recommendations on technical feasibility in section 6 of this submission.

JJ. See recommendations in section 11 of this submission regarding the range of enforcement options that a provider may consider. As recommended in part 11, this should not be prescriptive and a provider should, instead, be required to enforce its contractual rights in respect of the breach in an appropriate way that is reasonably proportionate to the extent of the harm to the online safety of Australians that may reasonably be expected to flow from the breach. Requirements to take enforcement action should be consistent with the action required by section 14 (subject to our recommendations in part 11 of this submission) and should be subject to the technical feasibility test.

KK. The requirements to achieve removal of all instances of material should be deleted.

18. Requirements to provide information ‘in service’: DIS Standard, various provisions including, for example, section 28 and RES Standard, various provisions including, for example, section 19(9), 26, 27 and 28

There are a range of obligations in the Standards on service providers to provide mandatory information and mechanisms/tools ‘in service’ as opposed to on a separate website.

18.1. **Issues with ‘in-service’ information requirements:** The practical implementation of these requirements is problematic for some services. As drafted, these requirements assume that services are provided on a website like platform but this is not invariably the case. For some services, such as apps, the configuration of the service does not allow for large amounts of information or mechanisms/tools to be provided ‘in service’. Sometimes, due to the layout/structure of apps or services, ‘in service’ information is not necessarily the most clear and obvious way for the information to be provided (as it can actually sometimes make it more difficult for the user to locate). Secondly, where a provider has a broad range of services, it seems more sensible to provide much of this in one centralised and clear location rather than a multitude of times or to, at least, link back to a centralised location, as this is where users would naturally look for such information. We think these requirements should be simplified so as to give providers discretion to determine the best method to meet the obligation is to provide users with ‘ accessible’ information/tools.

Specific recommendations

LL. We suggest that the Standards be amended so that the obligation is to provide users with certain information/tools in a manner ‘that is clear and accessible to end-users of the service in Australia’.

19. Expanded reporting: DIS Standard, Part 4 Division 3 and RES Standard, Part 4 Division 3

The reporting requirements for DIS and RES providers in both Standards are more extensive than the reporting requirements for service providers subject to the OSA Codes. This includes, for example:

a) increased requirements to share material related to risk assessments and risk profiles;
b) increased requirements to share information about development programs;
c) requirements to provide reports regarding technical feasibility;
d) increased requirements to notify eSafety of new features where there is any increase in risk – such reports are now required as soon as practicable ‘after making the decision’ to add the new feature or function and previous language confirming that no confidential information is required to be disclosed as part of this has been removed; and

e) increased compliance reporting including a range of new reporting metrics. This includes things like average monthly active users in Australia for some services, as well as certain take-down metrics, complaint numbers and details about the measures deployed pursuant to the Standards, including supporting detail and evidence.

A number of limitations previously built into the OSA Codes have been removed. For instance, a number of the provisions that set minimum time periods to respond to reporting requests have been removed.

19.1. **Issues with implementing requirements concerning reporting:** It is unclear why there is a need to significantly increase the reporting requirements for DIS and RES across a range of diverse service types over and above the reporting requirements imposed for other service categories under the OSA Codes. Many of the reporting requirements in section 37 and 36 are unnecessarily duplicative of the requirements imposed on providers under section 56 and 58 of the OSA which already provide mechanisms for eSafety to obtain relevant information about providers' actions in relation to class 1 materials, both on an ad hoc and periodic basis. For example, the Commissioner can, through the BOSE process, ask for information about the activities that providers are undertaking to detect and deter CSAM and pro-terror material on their service, including the types of solutions that they are deploying for that purpose. In addition, we note that proposed changes to the BOSE give power to the Commissioner to request some of these categories of information such as information about the number of Australian end-users. See also: our comments in section 10 of this submission regarding development programs.

Specific recommendations

MM. We suggest that reports be confined to:

a) the steps that the provider has taken, including measures and controls the provider has implemented, to comply with applicable minimum compliance measures in the Standard;
b) confirmation from the provider that the steps, measures and controls are appropriate, including reasonable supporting details and evidence;
c) where applicable, details of any limitations in terms of technical feasibility on the service or the provider to identify, assess or take action in respect of class 1A material and class 1B material as required by the Standard;
d) where applicable, details of the most recent risk assessment for the service, including about the plan and methodology; and
e) details of the volume of CSEM and pro-terror materials removed from the service.

NN. The reporting requirements should be amended to ensure that industry participants are given a minimum of 2 months notice to comply with a request for a report, consistent with the equivalent provisions in the OSA codes.

20. Complaints: DIS standard, section 40 and RES standard, section 40

Section 40(2) of the RES and DIS standard contains new requirements in respect of the handling of complaints 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.

The requirement to notify the complainant of 'the outcome' of its investigations is not sufficiently clear. This requirement should only apply where a complainant has provided contact information as many complaints are made anonymously and the ability for users to do this should be preserved.

Specific recommendations

OO. We suggest that sections 40(2) (b) be amended to require providers to notify complainants only of the decision and action the providers have taken in respect of the information to which the complaint relates, where the complainant has provided contact information.

Part 3: Issues specific to RES Standard

21. Duplicated risk assessment requirements for pre-assessed electronic services

Section 38(2) of the RES Standard appears to contain an erroneous reference to a 'closed communication relevant electronic service; a dating service; and a gaming service with communications functionality' which are pre-assessed relevant electronic services. These service categories are covered already by the requirements of section 37.

Specific recommendations

PP. The references in section 38(2) to 'closed communication relevant electronic service; a dating service; and a gaming service with communications functionality' should be removed as these services reporting obligations are specified in section 37.

22. ‘Primary functionality’ as test for distinguishing between different RES categories

The Standards incorporate the concept of ‘primary functionality’ into many of the definitions (i.e., a service will fall within the definition only if its ‘primary functionality’ is consistent with that set out in the definition). For example, a 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.

Similarly, a closed communication relevant electronic service is only a pre-assessed relevant electronic service if *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 has the target end-users’ contact details other 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.

22.1. **Issues with ‘functionality’ as test for distinguishing sub-categories of services:** How a service is categorised under the RES standards determines which sections of the standards are applicable to the service. It is therefore critical that service providers are able to easily determine the relevant service category. We assume that the primary functionality test was aimed at avoiding potential overlap between service sub-categories in this Standard. However, the primary functionality test included in the revised definitions is very difficult to apply and could result in services that were intended to be in these categories falling outside of relevant definitions. We note that a functionality test is not part of the definition of different service types in the OSA, although the purpose of a service is relevant for determining whether a service is a ‘social media service’. The functionality test could be a significant issue if not revised. In addition, the reporting obligations for RES services as drafted require providers to explain why their services fall within a particular sub-categories, based on these definitions.

22.2. Two examples highlight the difficulty of applying the ‘primary functionality’ test to categorise services. For example, a dating service is a RES if its primary functionality is ‘to solicit, offer, promote or provide access to dating, relationship, compatibility, matrimonial, social or romantic referral service’. However, it is unclear what ‘functionalities’ of a dating service would distinguish dating services from other services with similar functionalities, for example, from escort services. Both dating services and escort services may have the same primary functionalities such as, allowing users to create profiles with photographs and enabling registered users to communicate with each other. Similarly, a RES is now only a ‘closed communication’ relevant electronic service if its ‘primary functionality’ includes enabling end-users to ‘create a list of other end-users of the service’ and accessing and communicating with people on that list. We suggest that the purpose of a dating service, like social media services, is a better way of distinguishing the service from an escort service. However, arguably, the ability to 'create a list' is not the primary functionality of most RES, it is an ancillary function. For instance, a user of an email service can choose to email other individuals without adding them to their contacts – it is the communication capability that is the primary function of the service, not the ability to create a list.

Specific recommendations

QQ. We suggest that the test of 'primary functionality' for RES services be removed.

23. Open communication RES

The definition of 'open communication relevant service' includes an avoidance of doubt provision: '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'. The meaning and scope of this provision, however, is confusing. It is unclear if this wording is intended to result in services that are otherwise 'closed communication relevant electronic services' falling into this category. For instance, should this provision be interpreted to mean a RES will be a closed communication relevant electronic service solely by virtue of the fact that it enables an end-user to invite an individual to communicate using their contact details (even if the contact details were obtained by the user independently from the platform)? It is critical that the distinction between closed communication RES and open communication RES is clear

Specific recommendations

RR. We suggest that the avoidance of doubt provision in the definition of open communication services should be removed.

24. Tier 3 RES

This term is undefined in the Standard.

Specific recommendations

SS. We suggest that the Standard be amended to include a definition of a Tier 3 RES.

Part 4: Issues specific to DIS Standard

25. Operating systems

The Equipment Code of the OSA Codes applies to providers of operating systems: i.e., a designated internet service that consists of an operating system for an interactive (Tier 1) device that is provided to Australian end-users. The DIS standard needs to make clear that it does not apply to operating systems or to OS providers as defined in the Equipment Code for clarity.

Specific recommendations

TT. We suggest that the Standard be amended to make clear that operating systems, or to OS providers (to the extent providers are acting in that capacity), as defined in the Equipment Code are not subject to the Standard.

26. Classified DIS

The definition of classified DIS is confusing and makes reference to material being classified as 'Category 1 or lower'. We assume that the intention is to refer to material being classified as 'Category 1 or Category 2 Restricted under the Classification Guidelines for Publications'. If so, this should be made clear in the definition.

Specific recommendations

UU. We suggest that the definition of classified DIS make explicit reference to the relevant categories under the Classification Guidelines for Publications.

27. High impact generative AI DIS and machine learning model platforms.

The DIS standard introduces new requirements for 'High impact generative AI DIS and machine learning model platforms'. These are new categories of service that have been developed by eSafety and the intended scope and application of these requirements is unclear. This also gets ahead of a broader Australian government approach to regulating AI and could end up being conflicting or redundant as a new framework is developed.

27.1. **Issues with approach to these new categories of services:** The definition of machine learning model platform seems to be more properly treated as a hosting service, as that term is defined in the OSA. While it is likely that some machine learning models would individually fall to be treated as a DIS, we question whether platforms that host these services should be treated as DIS under this Standard.

The definition of high impact generative AI DIS is very broad and does not distinguish between the different ways in which generative AI capabilities are developed, deployed and embedded in services that are ultimately made available to end-users. As a general point, we suggest that many services would consider the implementation of AI capabilities as adding new functions to their services, rather than resulting in the creation of a new service. We suggest it would be preferable to regulate those capabilities, rather than creating separate service categories that overlap with other service categories, which adds to the complexity of online safety regulation and is potentially confusing. It is also very important that the way these services are regulated under the Standard takes into account the complex supply chain for these technologies, and that regulations are targeted at addressing the risk that materials will be generated at the point in the supply chain where it can best be mitigated. With regards to harmful content outputs, this would most likely be at the application layer of the stack.

A variety of different entities may be involved in the creation and deployment of generative AI capabilities that are embedded in services that are ultimately accessible by Australian end-users. These different entities will have different levels of control over the services' capability to generate material, depending on where they sit in the tech stack. For example, foundation models of generative AI are often distinguished from other artificial intelligence (AI) models, which may be designed for a specific purpose. The later models may by virtue of their design be at very low risk of being used in a service in a way that may generate class 1 materials. Foundation models can be defined as 'AI models designed to produce a wide and general variety of outputs'.[27] They can be standalone systems or can be used as a 'base' for many other applications'. In the case of foundation models there can be multiple developers and deployers of the model in the supply chain. Providers of generative AI foundation models such as Open AI's foundation model GPT-4 may allow the model to be accessed by other companies ('downstream' in the supply chain) to build AI applications 'on top' of a foundation model, using a local copy of a foundation model or an application programming interface (API). Additionally, the provider of a foundation model may allow downstream companies to create a customised version of the foundation model using a process called 'fine-tuning' – which describes when new data is added to a foundation model to 'fine-tune' its performance and capabilities on specific tasks.[28] For this reason, online safety standards are best targeted at the application layer of the tech stack rather than further downstream at the foundation model level, particularly in the context of the broader government consideration of AI that is underway.

Question 17 of the Discussion paper asks: The high impact generative AI designated internet services category only captures models that meet a high impact threshold. It

[27] Elliott Jones, *Explainer: What is a foundation model?* Accessible at: https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/
[28] Ibid.

must be reasonably foreseeable that a service can be used to generate synthetic high impact material that would be classified as X18+ or RC. Is this threshold:

(a) appropriate for differentiating high impact and therefore high risk models?

(b) sufficiently clear to enable service providers to assess whether or not they meet the definition?

We consider that the definition of ‘high impact generative AI designated internet services’ as drafted is not sufficiently clear. In particular, it is not clear when it is ‘reasonably foreseeable’ that a service will be used to generate high impact materials and how that relates to the likelihood the service will be used to generate class 1 materials which will then be used to cause harm to users. For example, does a service qualify as a ‘high-generative AI DIS’ where steps have been taken to prevent the service being used to generate ‘high-impact materials’. For example, where the provider of the underlying foundation model has implemented technical or contractual restrictions on the service which aim to prevent it being used to generate high impact materials.

Question 15 of the Discussion paper asks: eSafety is seeking to place requirements on service providers that are best-placed to prevent the use of generative AI features to create and disseminate class 1A and class 1B material. Does the proposal achieve this? Question 19 of the discussion paper additionally asks: In relation to machine learning model platform services, do the proposed obligations (in particular, the section 23 obligation to ‘disrupt and deter’) provide appropriate safeguards? Are there specific challenges to deploying these measures? In the case of high impact generative AI DIS, it appears that the company that makes the service available to ‘end-users' will be treated as the provider of the service. In the case of a machine learning platform, it appears that the provider of the platform will be treated as the provider of the service. In our view, it is questionable that these ‘providers’ are invariably best placed to prevent the use of generative AI features to create and disseminate class 1A and class 1B material. For example, section 23 requires providers to take steps to disrupt and deter use of the service to solicit, generate, access, distribute, store or otherwise make available CSAM and pro-terror material. However, the ultimate provider of the service or the provider of a machine learning platform may not, in practice, be best placed or indeed technically or legally able to comply with these requirements. Rather, it is at the application layer where most control exists. For example, a provider of a generative AI service may have limited control over the underlying foundation model but only the fine-tuning of its performance. Similarly, the provider of a machine learning platform may have limited visibility over how the machine learning models on the platform enable the generation of materials and limited control over how the model may be used by the deployer of the model, or used by end-users of a service on which the model is deployed.

Specific recommendations

VV. We suggest that eSafety remove the requirements for high impact generative AI DIS and machine learning platforms, for the time being, and focus on the application layer of the AI tech stack. We consider that the requirements for the regulation of these services will benefit from a more extended public consultation that enables eSafety to work through the complex issues raised by the proposed requirements for these service types and ensure that regulation takes place at the optimal point in the supply chain. This may best be achieved as part of the review of the Online Safety Act, which is scheduled to commence early in 2024.[29]

[29] Michelle Rowland MP, *Media Release*, 22 November 2023 accessible at: https://minister.infrastructure.gov.au/rowland/media-release/albanese-government-takes-major-steps-forward-improve-online-safety.