COMMUNICATIONS ALLIANCE LTD

Communications Alliance Submission

to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts

Online Safety (Basic Online Safety Expectations) Amendment Determination 2023

and ancillary materials:

Amending the *Online Safety (Basic Online Safety Expectations) Determination 2022* – Consultation paper

Amending the *Online Safety (Basic Online Safety Expectations) Determination 2022* (BOSE Determination)—Summary of the BOSE Determination and proposed amendments

16 February 2024

CONTENTS

Communications Alliance

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

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

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

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

Introduction

Communications Alliance welcomes the opportunity to make a submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (Department) in response to the *Online Safety (Basic Online Safety Expectations) Amendment Determination 2023* (Amendment), the associated Consultation Paper and Summary of the proposed amendments.

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

We welcome initiatives that aim to further online safety through a wholistic framework that aligns to and harmonises with other legislative and regulatory frameworks. However, we are concerned that the proposed amendment to the BOSE further focus the BOSE away from its original intent of transparency reporting, continue to be overly broad in their application to service providers and material in scope, and exacerbate existing issues by further widening the scope of material to be considered as part of the expectations and through inclusion of expectations for specific technologies.

Importantly, we are bewildered by the timing of the proposed Amendment given the multitude of processes already on foot that closely interrelate with many of the proposed amendments, in particular the statutory review of the *Online Safety Act 2021,* which forms the legislative underpinning of the BOSE, the development of the class 1 industry standards, the development of class 2 codes (or standards as the case may be) and the outcomes of the *Privacy Act 1988* review and foreshadowed co-design of AI regulation.

1. Timing of the proposed Amendment

1.1. As the Consultation Paper correctly indicates (pp. 4/5), the proposed Amendment of the *Basic Online Safety Expectations* (BOSE) occurs against the background of several other reform processes. Those include:

- the (early) independent statutory review of the legislation underlying the BOSE, i.e. the *Online Safety Act 2021* (Act);
- the making of industry standards for class 1 material for relevant electronic services (RES) and designated internet services (DIS) by the Office of the eSafety Commissioner (eSafety);
- the development of industry codes for class 2 material;
- activity flowing from Government's interim response to the *Safe and responsible AI in Australia* consultation, including the proposed AI Safety Standard to be co-designed with industry;
- the review of the *Privacy Act 1988*, including Government's agreement to implement a Children's Online Privacy Code to promote the design of certain services in the 'best interests of the child';[1]
- the *Misinformation and Disinformation Bill 2023* and associated processes;
- the voluntary code for online dating services;
- the Department of Home Affairs report in relation to understanding algorithms on digital platforms; and
- the anticipated Government response in relation to dispute and complaints resolution processes of digital platforms, flowing from the ACCC's *Digital Platform Inquiry*.

1.2. All of the above processes have the potential or are expressly intended to substantially influence operational and design aspects of the services in scope of the BOSE.

1.3. Importantly, the Department released the terms of reference for the statutory review of the Act on 13 February 2024. We take this opportunity to draw attention to a number of relevant elements of the terms of reference. Item 3 of the terms of reference identifies the operation and effectiveness of the BOSE as an area the review will consider. In addition, Item 4 of the terms of reference sets out additional arrangements not explicitly captured under the existing statutory schemes that the review should consider, including online hate, and potential online safety harms raised by a range of emerging technologies, such as generative artificial intelligence and recommender systems.

From a process and public policy perspective, it is difficult to understand the purpose of amending the BOSE at the same time as a comprehensive statutory review that will examine not only the operation and effectiveness of the BOSE, but also areas the amendments to the BOSE already propose to incorporate prior to the conclusion of the review (let alone revision) of the Act. These parallel processes could result in an outcome where the Amendment of the BOSE conflicts with the recommendations of the review's final report. Alternatively, one could argue that the Amendment presupposes certain outcomes of the review of the Act. This would be a poor outcome, and is likely to create unnecessary compliance complexities for services providers.

[1] p. 15, Australian Government, *Government Response Privacy Act Review Report*, 28 Sept 2023, as accessed at https://www.ag.gov.au/rights-and-protections/publications/government-response-privacy-act-review-report on 6 Feb 2024: "To clarify how the best interests of the child should be upheld in the design of online services, and provide further guidance on how entities are expected to meet requirements regarding targeting, direct marketing and trading, the Government **agrees** a Children's Online Privacy code should be developed (*proposal 16.5*) as soon as legislated protections for children are enacted to enable the development of such an APP code. The code would apply to online services that are likely to be accessed by children. To the extent possible, the scope of the code should align with international approaches, including the UK Age Appropriate Design Code, with similar exemptions for particular entities such as counselling services. The code developer should consult broadly with children, parents, child development experts, child welfare advocates and industry in developing the code."

1.4. In addition to the statutory review of the Act, some of the proposed amendments to the BOSE are likely to interact with one or more of the above processes, for example in relation to artificial intelligence (AI, also refer to our comments at item 5 below), the considerations of the 'best interests of the child' (refer to our comments at item 8) and the treatment of class 2 material.

1.5. A number of proposed amendments also substantially overlap with other parts of the online safety regime such as the industry codes of practice for class 1A and class 1B material, and the draft RES/DIS standards, thereby potentially creating confusion for participants, inconsistencies and unnecessary duplication. (Also refer to our comments at items 3 and 9 below.)

1.6. Consequently, industry fails to understand the rationale for amending the BOSE at this stage, i.e. prior to an already scheduled and soon to be completed review of the Act, prior to the finalisation of the industry standards, the development of class 2 codes and in parallel to other relevant processes noted above.

1.7. In our view, a potentially earlier implementation of the proposed amendments does not justify the likely drawbacks resulting from potential inconsistencies with other processes and/or the foregone benefits that could be achieved from the completion of at least some of the above processes (noted at item 1.1) prior to the making of the Amendment.

1.8. Further, the piecemeal nature of proposed AI regulation does not appear to be aligned with the approach Minister Husic has outlined[2], i. e. with AI regulation occurring only in high-risk settings and where clear gaps in existing regimes have been identified. Instead, we are seeing overlapping, iterative regulation and a lack of holistic consideration of the harms and gaps that are sought to be addressed, with limited evidence of consideration of the impact this will have on the development and take-up of AI in Australia.

1.9. We ask the Department to provide a detailed explanation as to the necessity of the Amendment at this time against the background of the processes already underway, in particular the review of the Act and the development of the class 2 codes, and urge the Department to re-consider the timing (as well as some of the detail) of the Amendment.

1.10. We also urge the Department, in close cooperation with other relevant Departments, to holistically consider the regulation of AI. We suggest that the further development of regulation touching on AI should be temporarily paused until the various items of regulation can be evaluated and reconciled and, subsequently, to align those streams with the review of the Act and the development of the AI Safety Standard.

1.11. Notwithstanding our recommendation that amendments to the BOSE should be considered once the statutory review of the Act has concluded, we make a number of observations on what the Department has proposed below.

2. Regulatory approach and purpose of the BOSE

Regulatory powers and approach

2.1. As currently applied in practice, the BOSE are far more than a transparency reporting regime.

2.2. The Act confers substantial powers onto eSafety in connection with the BOSE, including significant reporting obligations for providers in response to a notice from eSafety. This essentially allows eSafety to repeatedly compel information from providers at short

[2] Refer to Minister for Industry and Science, 17 Jan 2024, *Action to help ensure AI is safe and responsible* as accessed on 16 Feb 2024.

notice,. The Act also includes civil penalty provisions for non-compliance with such a notice. eSafety has the power to issue formal warnings, service provider notifications and infringement notifications, and can accept enforceable undertakings or seek injunctions. Importantly, eSafety can 'name and shame' providers that, in eSafety's view, are not meeting the expectations.

2.3. Further, expectation 7 of the BOSE anticipates that *every* social media service, RES and DIS available in Australia will consult with eSafety when determining 'reasonable steps' for the purposes of expectation 6(1) and will have regard to relevant guidance material from eSafety.

2.4. This places eSafety in a position of significant power under the Act with respect to the BOSE.

2.5. As a result, when coupled with the broad and undefined scope of the BOSE (see comments directly below from item 2.6), this leads to a position where eSafety's views on, and approach to, the BOSE is a key consideration when reviewing possible amendments to the regime.

Scope of the BOSE

2.6. Importantly, the BOSE are very broad – in our view overly broad – with respect to both the entities they cover and the scope of the material to which they apply.

- The BOSE apply to all social media service providers, RES and DIS providers, irrespective of size, risk profile or feasibility assessment. In effect, this means that the BOSE apply equally to the majority of websites and apps available in Australia. In our view, as more specificity is built into the BOSE in practice (see further from item 2.9 below), this is resulting in a heightened regulatory burden that is arguably not proportionate to the risk level posed by the extremely broad range of services in scope, and expectations that are not tailored in any reasonable way to different service types.
- The original policy intention of the BOSE, as expressed in the Explanatory Memorandum accompanying the Act, was to apply the BOSE narrowly, initially only to social media services. This is also evident from the regulation impact statement which only pertained to social media services.

  However, the BOSE as implemented apply equally to all social media services, RES and DIS. Further, the legislative drafting of the definitions for both RES and DIS are so broad and vague that most online services can find themselves within scope of the BOSE, without any proper consideration being given as to whether the requirements are proportionate, necessary and targeted for different service types, including services that are emerging and evolving.
- The material in scope not only includes material that is within the specific categories of content dealt with under the Act (which are expressly listed in section 46 of the Act as the object of the BOSE core expectations). Instead, the BOSE extend to all 'unlawful and harmful' material and activity, with a focus on use of services in 'a safe manner'– none of which is defined. Safety and harm are extremely broad concepts, and the lack of definitions gives rise to significant scoping concerns and overlap with other statutory regimes. This is an existing issue with the BOSE. However, the introduction of provisions regarding hate speech is a good example of how additional categories of content (including content normally the subject of other legislative regimes, or within the remit of other regulators) can be added into the scheme due to the lack of definition. See item 7 below for more detail on this point.

2.7. Accordingly, eSafety has wide discretion as to its assessment of whether it considers material or activity harmful or unsafe and, consequently, whether an expectation ought to have applied.

eSafety also, in practice, exercises considerable discretion over the specific questions it considers relevant for providers to answer with respect to each expectation (rather than asking providers to report on the steps taken to comply with the applicable basic online safety expectations during the reporting period).

Similarly, it is within eSafety's discretion to determine whether an expectation ought to have been met by a specific service provider, and there is no obligation to consider the level of risk a service provider or category of service provider poses. As we expand on below (at item 3), this is at odds with the risk-based approach taken in the industry codes and standards.

2.8. Noting that neither the BOSE nor the decisions taken by eSafety when it exercises its considerable discretion as to the best way to approach an undefined set of online 'harms' are subject to parliamentary debate, we believe that this blanket approach to the scope of entities and material and the resultant discretionary approach does not constitute good regulation and ought to be reconsidered in-principle.

Specificity of the BOSE

2.9. The BOSE were expected to operate as a set of broad and flexible statutory expectations underpinning a transparency reporting regime. The accompanying *Amending the Online Safety (Basic Online Safety Expectations) Determination 2022 (BOSE Determination)—Summary of the BOSE Determination and proposed amendments, Summary of the BOSE Determination and proposed amendments* for instance states that the

> *"BOSE Determination does not prescribe how these expectations will be met. Rather, they are drafted in a way that allows flexibility in the method of achieving these expectations".*[3]

2.10. The BOSE sit alongside industry codes and standards containing mandatory minimum obligations for providers. The BOSE and the industry codes and standards are therefore two separate but complementary regimes – they have quite different functions under the Act.

2.11. We are concerned that the Amendment further removes flexibility and that, in practice, the BOSE is being treated as a mechanism to impose quite specific obligations when that was not its function.

- Some of the new proposed expectations read as obligations, as opposed to expectations that allow providers flexibility as to how the expectation will be met. As such, these provisions inevitably cross over and conflict with similar requirements under the codes and standards. See items 3 and 9 below for more detail.
- In addition, when the level of specificity being added into the BOSE is combined with eSafety's own detailed expectations (as reflected in documents such as the *Basic Online Safety Regulatory Guidance*[4]), we believe that, in practice, providers

[3] p. 1, Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *Amending the Online Safety (Basic Online Safety Expectations) Determination 2022 (BOSE Determination)—Summary of the BOSE Determination and proposed amendments, Summary of the BOSE Determination and proposed amendments, Oct 2023*

[4] Whilst the Guidance indicates that providers "*may elect to take different steps [other than the examples] to meet the Expectations that better suit their services and the risks posed*" the Guidance nonetheless goes on to use language such as "*providers should*", "providers should never" or "*eSafety expects*" to discuss many of the examples.

are being expected to comply with very specific expectations regarding how the BOSE will be implemented. There is a significant risk that any new examples included in the regime may in practice be treated as mandatory, again creating inevitable cross-over and conflict with similar requirements in the codes and standards. The new examples as drafted are worded very broadly and do not take account of differences in risk or service type. Having these applied as mandatory in practice, would create significant problems for a broad range of services.

2.12. We also note with concern that the proposed Amendment includes additional expectations in relation to AI, recommender systems (and a new specific type of material, i.e. hate speech, as discussed further above).

2.13. The 2021 *Frequently Asked Questions – Basic Online Safety Expectations* expressly note that

> "*[t]The Expectations are drafted so as to avoid being overly prescriptive and specific.*"[5]

2.14. The inclusion of specific expectations in relation to specific technologies, such as AI and recommender systems is concerning for two reasons: First, it caters to the already existing tendency of the BOSE to contain a greater extent of specificity than what industry originally understood to be intended (as evidenced in the 2021 FAQ). Second, it risks creating expectations – and investments to meet those expectations – that may be premature or ill-targeted given the early stages of some technologies.

We do not believe that it is advisable to amend the BOSE for specific new (or existing) technologies but, rather, to recommend language for the expectations that is technology-neutral and outcomes-oriented.

2.15. It is inappropriate for this instrument to set technological or definitional precedents, i.e. the instrument is unsuitable to attempt to establish concepts, definitions or expectations for specific technologies (including new technologies such as generative AI) or categories of content that are not within the specific categories of content dealt with under the Act.

3. Interaction of the BOSE with the draft RES/DIS standards and registered codes

3.1. As previously indicated, the Amendment contains a number of additional expectations and examples that overlap – though do not completely align – with requirements of the draft RES/DIS standards and registered industry codes.

3.2. In many instances, the Amendment extends the requirements of the draft standards to all RES and DIS (extension of services in scope due to lack of risk tiers) while at the same time expanding the scope of the material covered from class 1A and 1B to 'unlawful and harmful' material including hate speech.

As a result, and due to the absence of a risk assessment process and/or definitional confinement of expectations to only sub-categories of services that would remove the expectations from some low risk services (as attempted in the draft standards), or refine the expectations as they apply to sub-categories of services (to address technical or other service-specific differences), a number of the proposed amendments are simply

This suggests strongly that eSafety's expectations may not necessarily be aligned with a provider choosing to take a different approach may result in examples provided in the BOSE being treated as part of the expectations in many instances.

[5] p. 2, Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *Frequently Asked Questions – Basic Online Safety Expectations*, Oct 2021 as accessed at https://www.infrastructure.gov.au/sites/default/files/documents/frequently-asked-questions--basic-online-safety-expectations.pdf, on 6 Feb 2024

unrealistic or impossible to be complied with by a very large number of services to which they apply, e.g. the majority of websites (DIS).

3.3. This overlap can be found with respect to requirements (standards/code) and expectations (BOSE) that relate to:

- generative AI;
- assessments and ensure safety by design;
- reporting and complaint mechanisms for end-users;
- investments in systems, processes and/or technologies to further improve capabilities; and
- terms of use, policies and procedures to address the relevant material, and enforcement of such terms, policies and/or procedures.

3.4. The following may serve as examples to provide further detail on the extent of overlap.

- proposed section 8A – regarding reasonable steps in relation to generative artificial intelligence capabilities:
  - → refer to draft DIS standard sections 21, 23 and 25;
- proposed paragraph 6(3)(f) regarding the assessment of the impact of business decisions:
  - → refer to sections 8 and 9 of the draft RES/DIS standards in relation to conducting risk assessments for material changes.
    Note, however, that some service categories are exempted from the risk assessment requirements (pre-defined/assessed services), thereby creating further confusion as to what is required and/or expected of service providers.
- proposed paragraph 6(3)(g) regarding actioning reports and complaints in a reasonable time:
  - → refer to sections 30 and 31 of the draft DIS standard and section 30 of the draft RES standard and other sections in relation to reports of breaches of terms of service.
- proposed paragraph 6(3)(g) regarding investments to improve the detection of material and activities:
  - → refer to the requirements around development programs in the draft RES/DIS standards.
- proposed section 18A to publish transparency reports:
  - → refer to the reporting obligations of the draft standards. It is likely that in order to discharge of the requirements contained in the draft standards, providers will need to provide information that overlaps with the information contained in the proposed new section. In the absence of any confidentiality requirements on the Commissioner, this information can be published by eSafety.
    For further commentary on the proposed transparency reports, also refer to our comments at item 9.

3.5. These overlaps create unnecessary duplication and complexity and, at worst, inconsistencies. As highlighted above, the resulting difficulties stem from, in our view, an inappropriate focus of the BOSE away from a transparency regime to a quasi-mandatory regime of additional requirements. Consequently, conflicts with the industry standards and registered codes are bound to arise. This could be remedied by refocusing the BOSE on its originally intended function of transparency reporting.

3.6. As currently drafted (and as noted in our submission in response to the draft standards), the draft RES/DIS standards are too complex to be understood and complied with by a large number of affected organisations. The interaction of the standards (and the registered codes) with the BOSE introduces additional complexity as it, in effect, imposes an additional and separate set of obligations for providers for the same type of content. The overlap highlighted above further exacerbates these difficulties.

3.7. Importantly, eSafety provided formal guidance that

> *"Compliance with the requirements in an industry code or industry standard is relevant to a provider's implementation of certain expectations (in relation to class 1 material) but will not be determinative of meeting any particular Expectation.*
>
> *This is because what is 'reasonable' for a provider to do to address unlawful and harmful material under the Expectations may extend beyond the minimum requirement in the mandatory (and enforceable) industry code or industry standard. Additional steps may be required to meet the applicable Expectations."*[6]

This guidance remains conceptually difficult – a difficulty that is exacerbated by the duplication of requirements and expectations: given the registered industry codes and standards (yet to be made) through virtue of their registration or making by eSafety, are being deemed to provide "*appropriate community safeguards*" (section 140 of the Act), it is not clear why an 'expectation' under the BOSE would extend beyond a 'minimum requirement' in a code/standard that has been deemed to meet "*appropriate community safeguards*" [emphasis added] even in cases where the service and the material are covered by both instruments.

Compliance with the codes/standards ought to amount to fulfilling the respective expectation in such cases. The guidance on the BOSE published by eSafety ought to be amended (by eSafety) accordingly.

However, in our view, it is Government's role to provide clarity as to the intended purpose of the BOSE in relation to the industry standards and registered codes, and to ensure that providers are not subject to duplicative, inconsistent and confusing obligations that unduly complicate compliance.

4. Control over technology and feasibility of proposals

4.1. With the exception of very few expectations (which target specific services), the current BOSE and the proposed Amendment apply to any social media service, RES or DIS.

4.2. As previously noted, we believe that the BOSE ought to be returned to their original intention of a flexible, broad set of expectations to underpin transparency reporting. To the extent that this is not achieved, we note the following:

4.3. Both versions of the BOSE lack a mechanism to limit expectations to those service providers for which compliance with the respective expectations is feasible, or to ensure expectations are proportionate to the risk size and maturity of the relevant provider.

4.4. A number of services, including RES provided by carriage services, encrypted services and many websites or apps, cannot comply with many of the stated expectations, either because compliance is not technically feasible, would require breach of other applicable legislation and/or would be disproportionate in terms of costs, privacy impacts and/or the likelihood of an unacceptable rate of 'false positives', when compared to the potential risk of harm emanating from a communication or service.

4.5. While substantially more work is required with respect to a feasibility concept in the RES/DIS standard, both the draft industry standards and the registered industry codes include a feasibility concept.

[6] p. 10, Office of the eSafety Commissioner, *Basic Online Safety Expectations Regulatory Guidance*, Sept 2023, as accessed at https://www.esafety.gov.au/sites/default/files/2023-09/Basic-Online-Safety-Expectations-Regulatory-Guidance-updated-September-2023_0.pdf on 6 Feb 2024

The BOSE also ought to include at a minimum such a concept – closely aligning with the registered industry codes – and ideally also include strong protections for security and privacy, as drafted in the registered industry codes.

4.6. While compliance with the expectations is not mandatory under the Act in the same way as is the case for codes and standards, eSafety has substantial powers to publish information in relation to a provider's actions with respect to the expectations (and to control the narrative) and to 'name and shame' providers for non-compliance with expectations, with potentially far-reaching reputational consequences for affected services providers.

Therefore, it is, in our view, not acceptable that the decision over whether or not a provider is capable of meeting an expectation or whether compliance is feasible is vested in eSafety without further guardrails in the regulation or legislation itself.

4.7. We recommend a re-assessment of each expectation (existing and proposed) against a catalogue of social media services, RES and DIS service categories (e.g. SMS/MMS, encrypted services, email (provided by carriage service provider and over-the-top providers), any website, any app etc). We also recommend a review of the expectations with view to services being provided to consumers or enterprise users. This will allow the Department to get a better understanding of the expectations that could realistically be met by the different service providers.

Subsequently, expectations ought to be differentiated by service category and only apply to those services where they are feasible and in a manner proportionate to the harm likely to emanate from a service.

5. AI

Inclusion of generative AI into the Amendment

5.1. We acknowledge Government's desire to address evolving risks associated with generative AI; while – hopefully – also acknowledging the potential benefits of this transformative technology.

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

5.3. In the Australian context, the Department of Industry, Science and Resources (DISR) is leading a whole-of-government process to develop regulatory mechanisms to ensure AI is used safely and responsibly. In addition, there are other related Australian Government initiatives, such as the review of Australia's privacy, cybersecurity and copyright regimes.

5.4. In our response to DISR's Discussion Paper *Safe and Responsible AI in Australia*, we articulated that to date, AI has been capably regulated through existing technology-neutral legislation and regulation.

We also stated that any gap analysis should take into account whether existing legislation or regulation was capable of appropriately assigning responsibility across the AI supply chain, and whether unlawful and illegal use of AI applications was appropriately prevented or mitigated by existing legislation or regulation.

As noted in the Consultation Paper to the Amendment and in the guidance to the current BOSE:

> *"The Expectations apply to material and activity that is unlawful or harmful, irrespective of how it is generated."*[7]

5.5. In mid-January 2024, DISR *released the Australian Government's interim response to the Safe and responsible AI in Australia* consultation. In its interim response, Government noted:

> *"While the government considers mandatory guardrails for AI development and use and next steps, it is also taking immediate action through:*
>
> - *working with industry to develop a voluntary AI Safety Standard, implementing risk-based guardrails for industry*
> - *working with industry to develop options for voluntary labelling and watermarking of AI-generated materials*
> - *establishing an expert advisory body to support the development of options for further AI guardrails."*[8]

The interim report further noted that

> *"[its work] will also consider links to other initiatives across the Australian Government as well as state and territory governments"*[9]

and highlighted the statutory review of the Act as one of the existing legal frameworks to be considered

> *"to ensure that the legislative framework remains responsive to online harms."*[10]

5.6. We agree with DISR's approach, i.e. that any regulation of generative AI – where it is deemed 'high-risk' – ought to follow careful regulatory co-design processes with industry and is based, to the largest extent possible, on existing legislation that has the capacity to address concerns in relation to the potential risks of AI more broadly.

5.7. It is critical that the Department's (and eSafety's) approach to addressing online safety risks posed by AI is consistent with the Australian Government's approach, as articulated in the interim response, as well as international standards. We are concerned that the proposed expectations (and subsequent compliance with those expectations) prematurely introduce concepts that do not align with developing global standards and yet-to be-made Australian regulation, and are difficult to understand and apply to real-world business models. We encourage the Department to engage with DISR as part of this process, and caution against the introduction of AI-specific language and expectations in the BOSE that is separate from this overarching regulatory process, thereby risking the creation of potentially contradictory or competing priorities.

5.8. The use of subordinate regulation in advance of the review of the underlying legislation – in this case the Act – is inappropriate and likely to cause inefficiencies and, worse, inconsistencies with other legislative frameworks that are not industry-specific.

5.9. We also observe that the language used in the Amendment in relation to generative AI is overly broad. For example, the preamble to section 8A(1) is extremely broad in stating that:

---

[7] p. 24, Office of the eSafety Commissioner, *Basic Online Safety Expectations Regulatory Guidance*, Sept 2023, as accessed at https://www.esafety.gov.au/sites/default/files/2023-09/Basic-Online-Safety-Expectations-Regulatory-Guidance-updated-September-2023_0.pdf on 6 Feb 2024

[8] p. 6, Department of Industry, Science and Research, *Safe and responsible AI in Australia consultation Australian Government's interim response,* Jan 2014 as accessed at https://consult.industry.gov.au/supporting-responsible-ai, on 8 Feb 2024

[9] p. 20, ibid

[10] p. 22, ibid

> *"If the service uses or enables the use of generative artificial intelligence capabilities, the provider of the service will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of artificial intelligence capabilities on the service."* [emphasis added]

The reference to 'use' or 'enabling the use' of generative AI would capture all internal uses of generative AI by service providers. It cannot be the intention that all internal uses of generative AI, such as generative AI features to assist with legal contract drafting or internal audit functions, are within scope of the BOSE. As noted above, we recommend language for the expectations that is technology-neutral and outcomes-oriented, and does not place blanket expectations on specific technologies.

Practical difficulties with the expectations in relation to generative AI on service providers as proposed

5.10. As already highlighted in our response to the draft RES/DIS standards, the proposed expectations equally fail to appropriately take into account the supply chain for AI technology and which entity in the supply chain has the control over the AI technology that would enable it to comply with the expectations.

5.11. Consider, for example, an enterprise DIS that makes an AI application available to its enterprise customers (e.g. Microsoft CoPilot). The enterprise DIS will not be able to control the application or the content generated as it cannot modify it. The ability to control the AI sits with the real provider of the AI as opposed to the organisation that made it available. The generation of content will be undertaken by the end-users of the customer organisation.

5.12. If the Department insists on including generative AI-specific expectations into the Amendment, we suggest the Amendment focus on the key differentiating factors, i.e. who controls or deploys the content to the end-user and has the relationship with the end-user.

Generally, only AI that is used by the end-user ought to be in scope (where feasible).

5.13. We also note that undefined and broad concepts such as 'harmful' content will cause additional difficulties in the context of generative AI.

6. Recommender systems

6.1. In addition to our general remarks with respect to the (in our view) unnecessary specificity of the BOSE (which is also evident in the inclusion of recommender systems) and the terms of reference of the Act (refer to item XX), we offer the following observations.

6.2. The proposed definition and expectations apply to recommender systems broadly, with recommender systems being defined (p. 8, Consultation Paper) as

> "*systems that prioritise content or make personalised content suggestions to users of online services. A key element of the system is the recommender algorithm, a set of computing instructions that determines what a user will be served based on [a range of] factors.*"

6.3. This definition does not contain any link to the risk of harm emanating from a recommender system or the recommendation itself. This could capture a vast array of services where users interact with goods, services or material online in a broad range of contexts, not just social media services or services providing access to user-generated material. Consequently, recommender systems that offer end-users, for example, alternative/additional suggestions for the purchase or use of goods and services

(including curated entertainment services) are also included in the scope of recommender systems that are to comply with the expectations, and subsequently are expected to provide evidence – and rely on eSafety accepting that evidence – that it would be disproportionate or infeasible for that system to comply with the expectation.

We believe the inclusion of recommender systems into the BOSE is impractical and unnecessary. If recommender systems must indeed be specifically addressed in the BOSE, then this ought to occur by tying the definition of such systems to a reasonable likelihood that a system causes serious harm (as defined in the Act) to an end-user. Otherwise, the BOSE would cast the net much more broadly than is necessary to address the relevant harm.

We would also suggest that the lack of reference to a harm concept means that the definition is fundamentally at odds with the government's stated approach to risk-based regulation of technology

6.4. We elaborate on our concerns with the application of the BOSE to 'harmful' content further below. However, irrespective of those concerns, it is key to note that, for the purpose of recommender systems, whether certain material is harmful is also likely to be a function of repeated exposure to (or recommendation of) the material in question and the reason why it was recommended. At what point and through what mechanisms would providers of recommender systems be expected to determine that the material was harmful to the end-user, rather serving an interest of a person to explore a certain topic in greater detail? While some analysis and subsequent action may be possible for some service types (and not necessarily in all circumstances), this is not the case for all services in scope.

6.5. Given the noted impracticalities, we recommend re-focusing the BOSE on the material to be addressed (i.e. unlawful and harmful material, however, note our comments with regard to a required linkage to the powers of the Commissioner) rather than individual technologies that may make such material accessible. This comment applies to material generated through AI and recommender systems.

7. Proposed scope of material

'Harmful' material and activity

7.1. The scope of the material covered by the BOSE is broad and undefined, and therefore subject to significant interpretation and subjectivity.

7.2. The current BOSE and the Amendment apply to 'unlawful and harmful' material and activity. While it is possible for providers to have a good understanding of what constitutes 'unlawful' material (this does not imply that it would necessarily be feasible, technically or otherwise, for a provider to detect such material), the same does not hold for the subjective standard as to what material would be considered 'harmful'. 'Harm' is a very broad concept, and the lack of definition gives rise to significant scoping concerns with the BOSE.

7.3. Given the central nature of the term 'harmful', we propose the term be defined for the purpose of the BOSE to mean material that is not illegal, but with regard to which the eSafety Commissioner has powers under the Act, i.e. cyber-bullying material, cyber abuse material and material in relation to intimate image-based abuse.

Beyond that, if Government believes that certain content is harmful such that it should not be permitted online, then this content ought to be made illegal, or, at the very least, it ought to be regulated in a transparent manner and process.

7.4. The lack of clarity (and resulting uncertainty) is also concerning for providers as section 221 of the Act specifically excludes liability for actions taken as a result of a remedial notice or a removal notice (and other notices) from the eSafety Commissioner,

however these exclusions of liability do not extend to actions taken voluntarily and in good faith by providers to proactively remove material from their services. In our view section 221 of the Act ought to be urgently amended to add relevant protections addressing the expanded requirements on service providers under the registered codes, industry standards and the BOSE.

7.5. The lack of clarity of what is expected is even more concerning given that providers are expected to ensure that end-users are able to use the service in a 'safe' manner – also an undefined term in the existing BOSE.

7.6. It is worth highlighting that the same issues of vagueness and subjectiveness described above are exacerbated through the vagueness of that term, i.e. it is unclear how 'safe' would be defined in this context. The concept of safety is subjective – while one user may feel safe, another user (of the same objective category of vulnerability) may not share that feeling.

7.7. However, we contend that even if the definition of 'harmful' was tied to the material to which the Commissioner's powers under the Act apply and which are expressly listed in section 46 of the Act as the object of the BOSE core expectations, it appears infeasible for providers of private communication services to determine, in the context of a private communication between two individuals, whether the material in the communication meets the (undefined or loose) criteria for cyber bullying or abuse, whether the material was shared consensually etc., without extensive knowledge of the context and background of that communication.

Hate speech

7.8. The proposed inclusion of hate speech into the scope of the material to be covered under the BOSE serves as a useful example as to how the additional categories of content – including content that is usually subject to other legislative regimes and/or substantial public debate (refer to our comments below) – can be added into the scheme due to the lack of definitional rigour, and despite the Act, so we believe, originally not intending the inclusion of such material in its scope.

7.9. The proposed extension of expectations to hate speech is very concerning for several reasons:

7.10. It is inappropriate to address hate speech in a piece of subordinate regulation that is not subject to parliamentary debate. Defining hate speech, balancing free expression rights, and determining a balanced approach to the issue in Australia is complex and difficult. For that very reason hate speech has, so far, not been addressed through comprehensive Commonwealth legislation.

7.11. In addition, State or Territory legislation that addresses components of hate speech – noting that in all circumstances such legislation provides substantially more definitional guidance than is proposed in the Amendment – and the *Racial Discrimination Act 1975* Cth confine the act of discrimination or hate to an act done in public.

Further, and similar to our comments in relation to harmful material in private communications, it would also be impossible to determine whether a communication constitutes hate speech in private communications.

7.12. It is inappropriate to attempt to address hate speech in a context limited to online material. This risks the creation of diverging approaches and standards to such speech and must be avoided. It is equally inappropriate to attempt to set a precedent for such material 'in the online world' through a subordinate regulatory instrument, irrespective of its 'voluntary' nature.

7.13. Section 474.13 of the *Criminal Code Act 1995* already makes the use of a carriage service to "*harass, menace or cause offence*" an offence. (Case law established (R v

Monis [2013]) that the degree of offensiveness must be serious. It must involve more than mere hurt feelings on the part of a reasonable person.)

7.14. Consequently, we request that any expectations specifically relating to hate speech be delayed pending the passing of the hate speech laws recently foreshadowed by the Attorney General – which are to be subject to parliamentary scrutiny – so that a coordinated and fully considered approach can be taken.

8. Best interests of the child

Interaction with other processes addressing children's rights online

8.1. We acknowledge that children ought to be afforded special protections. Members of our organisation already invest substantial resources in children's safety, in addition to general safety and cyber security measures.

8.2. We also welcome further regulatory and legislative work – in cooperation with all relevant stakeholders – in relation to children's rights online, including the foreshadowed changes to the *Privacy Act 1988* and the development of a *Children's Online Privacy Code*. We also note the terms of reference for the review of the Act which also include considerations of 'the best interests of the child' and a general 'duty of care'.

8.3. In this context, we are concerned about specific safety measures being proposed for implementation in subordinate legislation while overarching reviews of children's rights online are still taking place. We would like to see a coordinated approach to regulation, with these higher-level reviews completed and an overarching policy and legislative framework in place. Consequently, we believe it would be more useful to first complete or at least substantially progress this work prior to embarking on further children-specific amendments to the BOSE. This will ensure that the BOSE are consistent with and harmonise well with key economy-wide legislation and regulation.

8.4. Similarly, we note with concern that the proposed amendments specifically target the prevention of access to class 2 material and the development and improvement of technologies to prevent such access.

This does not appear useful, given industry associations will likely be asked, within the current quarter of 2024, to commence work on industry codes that, to a substantial extent, deal with access restrictions to class 2 material. Again, we believe that a better outcome will be achieved if those codes, that apply to all eight industry sections, are allowed to be developed prior to setting class 2-specific expectations in the BOSE.

Reference framework for the best interests of the child

8.5. If the Department insists on progressing the Amendment as proposed, we raise the following issues:

8.6. The current BOSE already contain an example (at section 6(3)(b)) to assist with compliance with the general obligation to minimise harm arising from material or activity on the service which requires that

> *"if a service or a component of a service (such as an online app or game) is targeted at, or being used by, children (the children's service)—ensuring that the default privacy and safety settings of the children's service are robust and set to the most restrictive level"*

8.7. The Amendment proposes an additional expectation at section 6(2A) to

> *"[…] take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any service that is used by, or accessible to, children."*

8.8. The Consultation Paper appears to aim at consistency of the new expectation with Article 3 of the *United Nations Convention of the Rights of the Child* (UNCRC) which states that

> *"In all actions concerning children, […], the best interests of the child shall be a primary consideration."*

8.9. Government, including eSafety, have previously also referenced[11] the UK Age Appropriate Design Code (AAD Code) as a useful model for Australian children's digital rights.

8.10. We agree that consistency with this UNCRC principle and the UK AAD Code forms a useful baseline also for children's rights online in Australia. Importantly, the AAP Code's explanation on the best interests of the child appropriately recognises that the best interests of the child, on the basis of the UNCRC, include considerations of a variety of needs that are to be considered in the best interests of the child alongside safety, including freedom of expression, privacy, agency to form their own views and have them heard, access to information, a right to association, play, etc. Without proper protections and proportionality, there is a risk that the BOSE could be interpreted in ways that disadvantage – or even infringe on – rights in the full suite of the UNCRC.

We strongly recommend that the Amendment clearly reference established guidance or establishes guidance consistent with these existing approaches to the best interests of the child to allow service providers to align with globally recognised and leading wholistic approaches to digital children's rights.

8.11. We agree with the understanding put forward in the Consultation Paper (and the AAP Code) that an analysis of the best interests of an individual child (and subsequent design considerations) is infeasible for most service providers with a large user base and child-users of different ages. This is even more so the case given the breadth of services in scope, which includes any service *"used by, or accessible to, children"*.

Consequently, it would be appropriate for the BOSE to clarify that service providers consider the best interests of the child users on their service more generally instead of the best interests of an individual child based on the particular circumstances of that child.

Services in scope for best interests of the child considerations

8.12. We have previously commented on the broad language used in this context, i.e. the expectation to apply the most restrictive default privacy and safety settings if the service is *"targeted at, or being used by, children"* [emphasis added]. Given children may randomly access or 'use' any app or service they can lay their hands on – in the vast majority of cases for a very short time and without any detriment as they are often not capable of meaningfully engaging with the service or app or quickly lose interest in it – the inclusion of mere 'use' in the expectation is not useful.

8.13. The proposed new expectation with respect to the best interests of the child is fraught with the same difficulty, however, exacerbates the already existing problem by now introducing even broader and deviating language, i.e. the best interests of the child are to be the primary consideration for *"any service that is used by, or accessible to, children"* [emphasis added].

[11] p. 152, Attorney-General's Department, *Privacy Act Review Report 2022*, Feb 2023 as accessed at https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report, on 9 Feb 2024

Almost any service is accessible to children and to make the best interests of the child a primary consideration in the design and operation will simply not be useful or feasible for a large number of services (noting that the BOSE apply to any website and app) and does not stand a proportionality or balancing-of-interest test.

Consider for example research websites or even the website and applications of the Australian Bureau of Statistics (ABS). It is clearly not useful to design those services and apps with the best interests of the child in mind. Given the very low risks of these services, other interests ought to be the primary consideration to design the best possible service for the targeted user group.

9. Transparency reporting

9.1. The Amendment proposes a new additional expectation for service providers to publish transparency reports. The expectation applies to all services in scope for the BOSE, i.e. all social media services, RES and DIS.

9.2. This contrasts the carefully balanced approach of the registered codes and the approach also forming the basis of the draft RES/DIS standards which seek to limit mandatory reporting (and regulatory burden – an objective shared by eSafety as expressed in the guidance to the BOSE) to those services where the risk of harm from the relevant material is highest. Other services are still required to provide reports on request.

We fail to understand how it could be feasible, or useful for that manner, to place a general public reporting expectation on any website accessible to Australians or any app available to them – how would providers of these services even know about the reporting requirements, let alone comply with them? Similarly, it does not appear useful to request public reporting from service providers for whom compliance with a range of expectations is infeasible and from whose services little risk of harm emanates.

As noted above, this approach, in our view, fails the test for good regulation as it creates unrealistic expectations that will also be impossible to enforce.

9.3. In addition, the proposed reporting expectation, if implemented, would create substantial overlap with mandatorily required reporting under the RES/DIS standard thereby leading to unnecessary duplication for similar but potentially not quite identical data sets.

9.4. Importantly, eSafety already has the requisite powers – and makes use of them – to request one-off and/or period reporting from service providers under the BOSE/Act. eSafety has also published requested information on a provider-specific basis.

9.5. The information to be reported on publicly as per the proposed Amendment may be confidential in nature. While such information may be reported to eSafety with respective requests for confidentiality, it is inappropriate to include potentially confidential information into a public reporting regime.

9.6. It is also not clear which benefit the public would derive with respect to minimising harm or protecting themselves from some of the information, such as the number of active users in Australia for a service (noting this may not be feasible to estimate, as discussed above, and may also require providers to disclose information that should be confidential for competition and other reasons). We submit that, in any case, the expectation for reports to include end user numbers should be removed.

9.7. In addition, the obligation to produce a transparency report per service, per country on a provider's enforcement efforts of its terms of service would imply that the provider is to report on all aspects of enforcement of the terms of service, irrespective of whether those efforts relate to the material covered by the BOSE (or Act) or other unrelated types of breaches, such as spam, scams, sharing malware, fraud etc.

9.8. Moreover, the expectation to provide information on a per-service, Australia-specific basis is, at best, impractical for some services. It may also not scalable, and would be an undue burden to report for a single specific market (within a set of global markets a provider may operate in) and it does not necessarily meaningfully add to or demonstrate how that service provider is tracking to address systemic harm across its systems.

The information is either not available in this form (potentially because it does not provide meaningful information or because it is technically infeasible) and/or it would require a high level of disproportionate effort to generate it, noting it requires production or segmentation of data in ways not currently generated or collated by providers.

9.9. On the basis of the above, we request that no additional reporting expectation be included into the BOSE as the public, through eSafety, already has access to information that could assist with minimising the experience of harm online.

9.10. Importantly, we also note that the new expectations relating to "*complaints and report handling*" (additional expectation 14(3) and the supplementary definitions in 14(4) and (14(5), the new proposed section 6(3)(g)) and "*enforcement of terms*" (additional expectation 14(1A) as well as amendments to sections 14(2) and 15(2)) all raise similar concerns as raised above in relation to transparency reporting.

We similarly request that these amendments also be reconsidered in light of overlap, inconsistency and necessity.

10. Miscellaneous and drafting issues

10.1. Our feedback provided in this section does not limit our previous comments.

Examples

10.2. With respect to examples listed in the BOSE, we reiterate that the BOSE ought to provide clarity (and eSafety consider providers' implementation of expectations in that light) that any examples in the BOSE as to how an expectation could be met are truly examples and will not be treated, in practice, as mandatory actions required in order to be considered fulfilling an expectation. This is particularly important as the examples are drafted in a broad manner that does not take account of service types and, consequently, examples sometimes are entirely unworkable for a range of providers in different instances. This is a lesser concern if the examples were truly recognised as such (i.e. examples) and providers were afforded flexibility as to the feasibility in the context of their own services.

It would be a significant concern if examples are, de facto or through discretion by eSafety, elevated to form part of the expectations.

User empowerment controls

10.3. Our feedback with respect to a lack of a feasibility test for services and expectations also applies to this additional expectation. Please refer to item 4.

10.4. We also note that the proposed subsection 6(6) appears to be intended as providing examples of reasonable steps that could be taken to meet the expectation, as is the case with other additional expectations. The Consultation Paper (p. 10) also refers to subsections 6(6)(a) to (c) as examples.

> *"(5) The provider of the service will take reasonable steps to make available controls that give end users the choice and autonomy to support safe online interactions.*

(6) Without limiting subsection (5), reasonable steps for the purposes of that subsection [could] include the following:

(a) making available blocking and muting controls for end-users;

(b) making available opt-in and opt-out measures regarding the types of content that end-users can receive;

(c) enabling end-users to make changes to their privacy and safety settings."

Subsection 6(6) is drafted such that (a) to (c) are actions that are all necessary in order to be deemed as having taken 'reasonable steps'.

We believe this to be an inadvertent drafting error and recommend correction by inserting a 'could' as highlighted in red.

Systems, processes and/or technologies

10.5. The Amendment refers to systems, tools, processes and technologies (with varying combinations of those terms) in a number of places.

Wherever these occur, they ought to be harmonised to 'systems, processes and/or technologies' [emphasis added] as it will not always be feasible for service providers to implement all of these measures.

This feedback corresponds with our feedback on the draft RES/DIS standard and aligns with the approach taken in the registered industry codes.

'Definition' of hate speech

10.6. A new subsection 6(4) attempts to 'define' hate speech as

"[…] a communication by an end-user that breaches a service's terms of use and, where applicable, breaches a service's policies and procedures or standards of conduct mentioned in section 14, and can include communication which expresses hate against a person or group of people on the basis of race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, gender identity, disease, immigrant status, asylum seeker or refugee status, or age."

While we understand the intention of the 'definition', the drafting would benefit from certainty and further clarification as not all communication by an end-user that breaches the terms of a service etc. will be hate speech, e.g. CSAM or pornographic material that are prohibited under a provider's terms of service would not constitute hate speech.

However, we reiterate our general rejection of the inclusion of hate speech into the BOSE.

11. Conclusion

Communications Alliance looks forward to continued engagement with the Department and all relevant stakeholders over ongoing reforms to Australia's legislative frameworks that govern online communications.

We urge the Department and other relevant stakeholders to reconsider the timing of the proposed Amendment as well as the key underlying principles of operation of the BOSE. We highlight our concern with any expectations that are overly broad, are not technology-neutral and/or run the risk of not aligning with other legislative or regulatory frameworks.

For any questions relating to this submission please contact Christiane Gillespie-Jones on [redacted]

Published by:
COMMUNICATIONS
ALLIANCE LTD

Level 12
75 Miller Street
North Sydney
NSW 2060 Australia

Correspondence
PO Box 444
Milsons Point
NSW 1565

T 61 2 9959 9111
F 61 2 9954 6136
E
info@commsalliance.com.au
www.commsalliance.com.au
ABN 56 078 026 507