To: Department of Infrastructure,Transport, Regional Development and Communication,
By email: BOSEreform@communications.gov.au

26 February 2024.

Dear BOSE reform team,

The Digital Industry Group Inc. (DIGI) thanks the Government for the opportunity to provide our views on the proposed amendments to the Online Safety (Basic Online Safety Expectations) Determination 2022 (BOSE) and accompanying Consultation Paper, which was released in November 2023.

DIGI shares the Government's commitment to improving online safety for Australians, which is a key pillar of our organisational mission, and agrees that online safety must be advanced at the platform level through safety by design, a core principle that underlies a range of regulatory instruments under the OSA, including the Class 1 industry codes for class 1A and class 1B material, proposed industry standards for relevant electronic services and designated internet services (also for class 1A and class 1B material) and the BOSE. In line with this commitment to online safety by DIGI and its members, we support the Government's objective of improving the BOSE by addressing issues and gaps in the original BOSE Determination, and of improving its overall operation.

DIGI shares the Government's commitment to improving online safety for Australians, which is a key pillar of our organisational mission. We agree that online safety must be advanced at the platform level through safety by design, a core principle that underlies a range of regulatory instruments under the OSA including the Class 1 industry codes, proposed industry standards for relevant electronic services and designated internet services and the BOSE. In line with this commitment to online safety by DIGI and its members, we support the Government's objective of improving the BOSE by addressing issues and gaps in the original BOSE Determination, and improving its overall operation.

At the outset, we note that the BOSE sets out basic online safety expectations for all social media services, relevant electronic services and designated internet services.[1] The BOSE is intended to be a flexible regulatory instrument that provides a broad set of common principles applicable to the sections of the industry that are in scope. They are one part of a comprehensive, overall regulatory framework for online safety in Australia. We note that – as with many parts of the online safety framework – the BOSE is effectively still in implementation: companies have only received no more than one round of non-period notices, and periodic notices have not yet been developed or issued.

[1] *Explanatory Statement, Online Safety Act 2021, Online Safety (Basic Online Safety Expectations) Determination* 2022, p.1.

Until now the BOSE, like the OSA, has been technologically neutral, focusing on the regulation of harmful material and activity rather than technologies. DIGI's principle aim in proposing the amendments below is to ensure that the approach of the BOSE continues to operate in a way that is compatible with the original policy intention and the broader OSA legislative framework. We suggest that changes to the BOSE that aim to expand the BOSE to deal with specific categories of harm (especially online hate speech as a standalone harm) and emerging technologies such as AI and recommender systems should be considered in the context of the forthcoming Statutory Review of the Online Safety Act, to ensure the online safety regulatory framework as a whole is clear, coherent and consistent.

We thank you for your consideration of the matters raised in this submission. Should you have any questions, please do not hesitate to contact me.

Best regards,

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

Table of contents

A. Need for a consistent approach to the administration of the BOSE, Codes and Standards

The BOSE are one part of a comprehensive, overall regulatory framework for online safety in Australia, and it is important to ensure the online safety regulatory framework as a whole is clear, coherent and consistent. In this section we set out DIGI's concerns about the effect of various inconsistencies between the BOSE, the OSA Codes and the proposed OSA Standards, and the need to ensure the BOSE is administered in a manner that requires the Commissioner to take into account considerations of proportionality and technical feasibility when considering whether the steps taken by the industry to meet specific expectations are reasonable.

1. Different standards of compliance for BOSE, Codes and Standards.

1.1. It is important that the BOSE amendments are compatible with the overall legislative framework established by the OSA, including the instruments contemplated by Part 9 of the OSA i.e. the industry OSA Codes and Standards. DIGI has a threshold concern that the proposed amendments to the BOSE overlap with and are in some cases inconsistent with the requirements of the Class 1 OSA Code for social media services which came into force on 16 December 2023 and the proposed Class 1 Standards for relevant electronic services and designated internet services (yet to be finalised). This inconsistency creates considerable confusion for providers of social media services, relevant electronic services and designated internet services about the steps that they should take to comply with these three instruments to the extent that they apply to the same

subject-matter i.e class 1A and class 1B materials. We consider that there is a need to clarify the relationship between these three instruments.

1.2. DIGI, alongside Communications Alliance, played a leading role between 2021 and 2023 in the drafting of the eight codes of practice under the OSA (Class1 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. In 2023, eSafety registered six of the Class 1 OSA Codes developed by industry including the Code developed for social media services (SMS Code). In 2023 eSafety also commenced a public consultation on proposed standards for relevant electronic services and designated internet services, which also apply to class 1A and class IB materials (RES/DIS Standards). As a result of our role in the Code development process, DIGI has unique insights concerning the interaction of the BOSE with the SMS Code and the proposed RES/DIS Standards.

1.3. The *Draft Online Safety (Basic Online Safety Expectations) Determination 2021 Consultation Paper* published in July 2021 (BOSE Consultation paper) made clear that the BOSE are intended to operate in a manner that is different to, but complementary with, the Codes and Standards. The BOSE Consultation paper explains that the BOSE is a flexible instrument which sets broad based expectations for a range of harmful activities and illegal and harmful materials (including class 1 and class 2 materials). In contrast, the Codes and Standards are intended to contain more specific minimum mandatory requirements for class 1 and class 2 materials:

> *The purpose of the Expectations is to place greater responsibility on service providers to ensure they provide safer services to Australian end-users. The Expectations provide flexibility for service providers to meet these Expectations. This approach recognises that traditional regulation may not suit the way content is created and delivered to users today. An example of this flexible approach is the expectation that service providers do more to assess and anticipate risks of harm facilitated by their services and take proactive and preventative action or 'reasonable steps' to mitigate those risks. Service providers are required to report on their compliance with the Expectations.*
>
> *The purpose of industry codes and standards is to set out binding self-regulatory procedures directed at ensuring class 1 and class 2 material is limited on services accessible to Australian end-users.*[2]

1.4. The proposed amendments to the BOSE include a range of new provisions that overlap with, but are somewhat inconsistent with, requirements of the SMS Code and the draft RES/DIS Standards. Examples of this overlap can be found with respect to requirements in the SMS Code and RES/DIS Standards and provisions in the BOSE that relate to:

- Generative AI;
- Risk assessments;

[2] Department of Infrastructure,Transport, Regional Development and Communications, *Draft Online Safety (Basic Online Safety Expectations) Determination 2021 Consultation Paper July 2021* p.5

- Complaint mechanisms for end-users;
- Investing in systems, tools and processes to improve the prevention and detection of material or activity on the service;
- Proactive detection of materials;
- Enforcement of terms, policies and/or procedures; and
- Transparency reporting.

1.5. The inconsistency between the three instruments means that where there is an overlap, it is now very unclear whether a provider of a SMS, DIS or RES service who complies with the relevant provisions of the applicable SMS Code or RES/DIS Standards will satisfy the requirements of the BOSE. This confusion is exacerbated by the *Basic Online Safety Expectations Regulatory Guidance* issued by the eSafety Commissioner in September 2023 (BOSE Guidance) which states 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.*[3]

1.6. We note that there is nothing in the OSA, the BOSE, or the Explanatory Memorandum for the BOSE that suggests that the standard of compliance for the BOSE in relation to class 1 and class 2 materials is higher or different to the standard required by the SMS Code and the DIS/RES Standards. Under the OSA, the suite of registered Class 1 OSA Codes and the DIS/RES Standards (when finally implemented) set 'appropriate community safeguards' in relation to Class 1 materials. Logically, compliance with the SMS Code and RES/DIS Standards should satisfy the equivalent requirements in the BOSE.

1.7. We recommend that the Government request the eSafety Commissioner to amend the BOSE Guidance to make clear that compliance with a Code or Standard under Part 9 of the OSA will satisfy the requirements of the BOSE to the extent that the Codes/Standards impose equivalent requirements on relevant providers with respect to Class 1 and Class 2 materials.

1.8. More generally, we note that there is inevitably going to be overlap and conflict if the OSA contains two separate, but different, mandatory or quasi-mandatory schemes aimed at exactly the same thing (at least to the extent that both cover class 1 and class 2 material). We therefore recommend that the different function and role of the BOSE under the OSA be carefully defined and maintained to avoid this. The BOSE were intended to be flexible statutory expectations underpinned by transparency reporting – but are increasingly being treated in practice as specific and mandatory or quasi-mandatory obligations (see, for example, the eSafety regulatory guidance on the BOSE which uses mandatory language to discuss a range of expectations and compliance examples). We

[3] Office of the eSafety Commissioner, *Basic Online Safety Expectations Regulatory Guidance*, Sept 2023 p.10.

suggest that the Codes and Standards are the appropriate regulatory scheme for minimum mandatory obligations aimed at high-impact content. We recommend that the BOSE be carefully defined and maintained in practice as a flexible set of expectations that providers are required to report against, to avoid inevitable inconsistency with Codes and Standards.

2. Need for consultation on BOSE Guidance.

2.1. Sub-section 7(1) of the BOSE requires providers of social media, DIS and RES services to consult with the Commissioner on what are reasonable steps for the purpose of meeting the core expectation in sub-section 6(1). Additionally, providers are obliged under sub-section 7(2) to have regard to any relevant guidance material published by eSafety. The Commissioner interprets the power to provide regulatory guidance under sub-section 7(2) very broadly. According to eSafety: 'This guidance may be updated in the future where additional guidance is required in relation to new harms, technologies and safety issues or in response to other events.'[4] The current BOSE guidance is quite detailed and includes guidance about a range of illegal materials and activity and emerging harms. It also includes a significant list of examples of reasonable steps for different expectations. In particular the guidance includes a list of interventions for recommender systems which are said to apply more broadly in relation to ensuring safe use of a service (sub-section 6(1) and 6(2).[5]

2.2. While we support the development of regulatory guidance which is updated to deal with new harms, it is important that industry has an opportunity to provide feedback before the guidance is released, so that the Commissioner is able to take industry's views into consideration, particularly where the guidance is extended to new technologies and emerging harms. The current BOSE Guidance was issued by eSafety without consultation with industry although when the BOSE was first released the Department advised industry that this would occur.[6] We will be raising this concern in the context of the forthcoming Statutory Review of the Online Safety Act. Alternatively, the Government may consider asking the eSafety Commissioner to provide for an appropriate period of industry consultation when the guidance is updated.

3. Need for Commissioner to take into account differences amongst providers including considerations of technical feasibility.

3.1. DIGI welcomes the eSafety Commissioner's statement in the BOSE guidance that recognises 'that differences between providers in terms of resources, risk, technical architecture and user base, means that 'one size does not fit all'.[7] A similar statement was made by the eSafety Commissioner in the *Basic Online Safety Expectations: Summary of industry responses to the first mandatory transparency notices published in December 2022*:

---

[4] Ibid p.33.
[5] Ibid p.31.
[6] Department of Infrastructure,Transport, Regional Development and Communications, *Frequently Asked Questions Basic Online Safety Expectations October 2021*,p.2.
[7] *Basic Online Safety Expectations Regulatory Guidance*, Sept 2023, p.13.

> *eSafety recognises that each provider is different – with different architectures, business models and user bases. This means an intervention or tool which may be proportionate and appropriate on one service, may not be on another.*[8]

3.2. While these statements by the Commissioner are very helpful, they are not embedded in the BOSE. As noted in paragraph 1.8 above, we are concerned to ensure that in practice the BOSE is administered in a flexible manner that permits the broad range of SMS, RES and DIS to adopt reasonable steps for their service type. We consider that in considering what steps are reasonable for their specific service, providers must be able to take into account whether specific types of tools and interventions (including those set out by eSafety in its regulatory guidance) are proportionate and technically feasible.

3.3. The principle of proportionality has been incorporated into section 5.1(b) of the Head Terms of the OSA Codes which include provision for providers of regulated products and services to implement measures in a manner that is proportionate to their 'size/scale and maturity' as well as their 'capacity and capabilities'. In the course of our work developing the draft DIS and RES Codes it became clear that, additionally, the services that fall within these categories had different degrees of legal and technical capability to comply with regulatory requirements concerning materials on their services. The draft RES and DIS Codes contained provisions that qualified the obligations of providers to implement certain measures when they did not have the legal or technical capability to do so. This concern has been accepted by eSafety and has now been carried over in the drafting of the RES/DIS Standards. In effect, the Standards exempt providers from complying with certain requirements, notably concerning proactive detection of certain types of class 1 materials where it is not 'technically feasible' to do so. We note that our submission on the RES/DIS Standards outlined a range of issues with how the technically feasible exception is currently applied in the draft RES/DIS standard, including that these exemptions should be further extended to a range of other requirements, for example those relating to complaint handling and enforcement of terms and conditions. We also noted that the concept of technical feasibility should be broadened to take account of other legal and practical constraints on various service types. For example, we have suggested that the concept should extend to prohibitions in telecommunications laws and interception and surveillance laws, technical and practical constraints posed by certain types of security measures and encryption, as well as constraints on providers dealing with end-users who are not their customers (e.g. in the case of email communications involving multiple email services).

3.4. The need to ensure that providers are able to take into account considerations of proportionality and flexibility is equally relevant to the BOSE. To take one example, the BOSE amendments introduce new expectations and examples of reasonable steps in sub-sections 14(1A), 14(2) and 15 around proactive detection of harmful and illegal materials and enforcement of a service's terms of use, policies and procedures end-users. However, some providers of relevant electronic services and designated internet services will have limited capability to meet these expectations or take the suggested reasonable steps for those expectations. For example, providers of encrypted services may have limited ability to assess communications in order to determine

[8] Office of the eSafety Commissioner, *Basic Online Safety Expectations: Summary of industry responses to the first mandatory transparency notices published in December 2022,* p.2.

whether terms of use have been breached or to proactively detect or remove materials. Critically, the terms of use for various platforms are wide ranging and clarification is needed around the feasibility of proactive detection on all violations of terms of use (which may include violations outside the scope of illegal materials and activities covered by the BOSE). We think that there is a need for the Government to give consideration to including further guardrails either in the BOSE or the OSA that make clear that the Commissioner must have regard to considerations of proportionality and technical feasibility (in the broad sense provided in the Class 1 Codes) when determining whether a service has taken reasonable steps to meet specific expectations.[9]

3.5. Given the power granted to eSafety with respect to the BOSE, the views of eSafety regarding the application and enforcement of the BOSE are a central consideration here. We support an approach that provides greater certainty for companies, as well as regular reviews of the operation of the BOSE, to ensure it is meeting policy expectations.

Recommendations

A. The Government request that the eSafety Commissioner amend the BOSE Guidance to make clear that compliance with a Code or Standard will satisfy the BOSE to the extent that they impose equivalent requirements on providers with respect to Class 1 and Class 2 materials.

B. The Government give consideration to amending the BOSE or the OSA that makes clear that the Commissioner must, when determining whether a provider has taken reasonable steps to meet all or any of the basic online safety expectations, take into account:

   a. the nature of the service;

   b. the context in which the service operates; and

   c. whether it is technically feasible (in the broad sense) for a provider to comply with particular expectations.

C. The Government request that the eSafety Commissioner provide for an appropriate period of industry consultation on the BOSE guidance.

D. The Government take steps to ensure that the BOSE operate as a flexible set of statutory expectations underpinned by appropriate transparency reporting.

[9] eSafety would for example need to make this assessment when issuing a 'service provider notification' identifying whether a provider has implemented the Expectations. These are also known as 'statements of compliance' and 'statements of non-compliance' - ss 55, 66 of the OSA.

B. Response to specific amendments

This section sets out DIGI's views and specific suggestions for improvement to the proposed amendments to the BOSE.

4. Additional expectation concerning best interests of the child

4.1. DIGI supports in principle the introduction of an additional expectation in subsection 6(2 A) of the BOSE concerning the best interest of the child, consistent with Article 3 of the UN Convention on the Rights of the Child (UNCRC).We have suggested a slight adjustment of the wording in the box below, to ensure the obligation is targeted at those services that are 'likely to be accessed by children' and not all services, some of which children are unlikely to access. We understand that 'the best interests of the child' also covers the full range of a child's rights, including a right to safety, to privacy, to free expression, to access information, to autonomy, and to a range of other critical rights.

4.2. We note that this is an area where we would appreciate additional regulatory guidance from the Commissioner concerning the steps that may be taken by providers to meet this expectation, noting that the guidance should be consistent with the international understanding of the best interests of the child concept. We understand that the Statutory review of the OSA will further consider this issue.

4.3. We note that the best interests of the child is one of the general principles contained in General comment No. 25 (2021) on children's rights in relation to the digital environment as formally adopted by the United Nations Committee on the Rights of the Child in February (2021).[10] The eSafety Commissioner supports this instrument, which also includes a range of additional rights of children which are pertinent to the BOSE and to the OSA more generally. We note that the Class 1 Codes developed by industry under the OSA impose obligations on providers subject to the Codes to consider 'the rights and best interests of children' more generally when implementing the relevant measures. This ensures that providers give consideration to the full range of childrens' rights online as set out in General comment No 25. We suggest that the Government give additional consideration to ensuring the expectation is extended to include all these rights consistent with the approach of the OSA Codes.

Recommendations

E. We suggest that proposed new expectation in subsection 6(2A) be amended to read as follows:

> The provider of the service will 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 *likely to be* accessed by children.

[10] Ibid

F. Consideration should be given to expanding the scope of the new expectation in subsection 6(2A) as follows:

> The provider of the service will take reasonable steps to ensure *that the rights of children in relation to the digital environmen*t including the best interests of the child are a primary consideration in the design and operation of any service that is used by, or *likely to be* accessed by, children.

5. Reasonable steps to ensure services can be used in a safe manner: business decisions

5.1. The proposed amendments include changes that provide additional examples of the reasonable steps that providers of services 'could' take to meet the core expectations in Section 6(1) and (2), including a new 6(3)(e):

> *ensuring that assessments of safety risks and impacts are undertaken, identified risks are appropriately mitigated, and safety review processes are implemented, throughout the design, development, deployment and post deployment stages for the service.*

DIGI agrees that safety is an integral aspect of good product governance for providers of digital services. We support the inclusion of the new proposed wording in 6(3)(e) which incentivises businesses to embed safety considerations within their risk management systems and processes for product development and ongoing management. However, we question whether anything is to be gained by the inclusion of sub-section 6(3)(f) which suggests as a reasonable step *'assessing whether business decisions will have a significant adverse impact on the ability of end-users to use the service in a safe manner and in such circumstances, appropriately mitigating the impact'*. We consider that the assessment of risk in relation to specific products is already covered by the new sub-section 6(3)(e). We note that sub-section 6(3)(f) is only provided as an example step that businesses should take to comply with expectations in sub-section 6(1) and (2). We assume that sub-section 6(3)(f) is intended to cover business decisions that are not specific to the development of particular products and services such as decisions on staffing, budgeting, expenditure, but the way that the provision is drafted implies that these types of decisions are likely to inherently incompatible with user safety, which is questionable, since they may equally serve to improve safety. We suggest that consideration be given to removing sub-section 6(3)(f) or reframing it in a more positive way (see below).

Recommendation

G. We suggest that proposed new expectation in subsection 6(3)(f) be amended to read as follows:

assessing how business decisions can support the ability of end-users to use the service in a safe manner.

6. Reasonable steps to ensure services can be used in a safe manner: Having processes for detecting and addressing hate speech.

6.1. The proposed amendments include as an additional examples of the reasonable steps that providers of services 'could' take to meet the core expectations in sub-section 6(1) and (2), a new sub-section 6(3)(i):

> *having processes for detecting and addressing hate speech which breaches a service's terms of use and, where applicable, breaches a service's policies and procedures and standards of conduct mentioned in section 14.*

Additionally, 'hate speech' is proposed to be defined in the BOSE as:

> *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.*

6.2. DIGI considers that this approach to hate speech proposed by this amendment is problematic for a number of reasons.

i. As drafted, it appears the regulatory intention is to introduce requirements on providers to deal with hate speech as a stand-alone harm. However, the OSA does not currently treat hate speech as a stand-alone harm. Insofar as hate speech is regulated by the OSA, it is treated as an individual harm in the form of cyberbullying. The inclusion of hate speech expressly in the BOSE highlights concerns with the unclear scope of the BOSE – in particular, its purported application to an indefinite range of illegal and harmful material and activities online as opposed to materials and activity specifically regulated by the OSA (as set out in section 46(1)(c) of the OSA).

ii. DIGI is concerned that there remains no clear legal standard nor recourse for victims of hate speech to incentivise strong action industry-wide. Notably, Australia's federal discrimination laws (as related to hate speech) focus on race and ethnic origin and are limited to acts that take place in public. The current laws do not cover hate speech related to other protected categories such as sexuality, gender identity, and disability. We do not think that the regulation of hate speech is advanced by a minor amendment to a subsidiary legislative instrument. In our view, the appropriate way to regulate hate speech is via

comprehensive laws of general application that apply both online and offline behaviours which are fully scrutinised and openly debated in parliament.

iii. DIGI has previously, and continues to, encourage the Australian Government to develop a clearer legislative framework that encompasses this broader approach to hate speech, and welcomes the inclusion of hate speech as a matter that will be considered in the context of the OSA review. We consider the OSA review to be an appropriate forum to consider whether changes should be made to the OSA or other legislative frameworks to address hate speech.The proposed amendment to the BOSE arguably constrains the independent reviewer's consideration of the issue in the context of the Statutory Review of the Online Safety Act, particularly consideration of what communications are properly regulated as 'hate speech' and the extent to which private communications on RES such as email and messaging services should be in scope, noting that provision concerning racial hatred in section 18C of the Racial Discrimination Act 1975 only applies to public behaviors. It is inappropriate to try to deal with the issue by a very minor amendment to in the BOSE before that review has concluded and the Government has had an opportunity to consider relevant recommendations, including any recommendations that address the fundamental questions of how hate speech should be defined and whether regulation should be limited to public behaviours.

Recommendation

H. We suggest that proposed new sub-section 6(3)(i) be deferred pending further consideration by the government of a clearer legislative framework that encompasses this broader approach to hate speech. DIGI supports the inclusion of hate speech in the Terms of Reference – Statutory Review of the Online Safety Act 2021 which is scheduled to conclude in October 2024.

7. New expectation concerning generative artificial intelligence capabilities.

7.1. A new sub-section 8A contains two new expectations that requires providers of *services that uses or enables the use of generative artificial intelligence capabilities to*:

1) *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.*

2) *take reasonable steps to proactively minimise the extent to which generative artificial intelligence capabilities may be used to produce material or facilitate activity that is unlawful or harmful.*

7.2. The OSA was drafted as a technologically neutral instrument, which was focused on protecting end-users in Australia from harmful materials and activities and would be able to adapt to the rapid technological changes that are characteristic of the digital

environment. The provisions of the OSA, as currently drafted, encompass harmful content and activity that is created by artificial intelligence. To date, the BOSE has also followed a technologically neutral approach. As noted in the Consultation Paper to the Amendment and in the BOSE Guidance: 'The Expectations apply to material and activity that is unlawful or harmful, irrespective of how it is generated.'[11] Providers that use or enable the use of artificial intelligence on their services are subject to all of the expectations, including the core obligation set out in the sub-section 6(1) of the BOSE that providers will take reasonable steps to ensure that end-users are able to use the service in a safe manner and additional core expectation in subsection 6(2) that 'the provider of the service will take reasonable steps to proactively minimise the extent to which material or activity on the service is unlawful or harmful'.

7.3. We understand that the proposed sub-section 8A has been introduced by the Government to deal with emerging risks with generative artificial intelligence[12]. However, as worded, sub-section 8A largely duplicates the expectations in sub-sections 6(1) and 6(2).

7.4. It is questionable whether a specific expectation around artificial intelligence is necessary given the broad scope of those expectations, or if it is appropriate given that it would be a considerable departure from the existing 'technology neutral' policy settings of the OSA. We also question whether the introduction of a new expectation focused on AI is consistent with the programme of work on AI being led by the Department of Industry, Science and Resources (DISR) to develop regulatory mechanisms to ensure that AI is used safely and responsibly. We understand that the Government's policy response on AI safety is articulated in the *Australian Government's interim response to the Safe and responsible AI in Australia consultation*, released by DISR in mid-January 2024. In the 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."*[13]

7.5. The Interim Report highlighted the statutory review of the OSA as one of the existing legal frameworks to be considered 'to ensure that the legislative framework remains responsive to online harms'.[14] Following the Interim Report, we understand that the forthcoming Statutory Review of the Online Safety Act will encompass 'other potential

---

[11] Office of the eSafety Commissioner, *Basic Online Safety Expectations Regulatory Guidance*, September 2023, p. 24.

[12] Department of Infrastructure,Transport, Regional Development and Communication, *Amending the Online Safety (Basic Online Safety Expectations) Determination 2022 (BOSE Determination*) Consultation Paper p.3

[13] Department of Industry, Science and Research, *Safe and responsible AI in Australia consultation Australian Government's interim response,* Jan 2024, accessed at https://consult.industry.gov.au/supporting-responsible-ai, on 15 Feb 2024, p.6

[14] Ibid, p.22

online safety harms raised by a *range of emerging technologies*, including but not limited to generative artificial intelligence, immersive technologies recommender systems, end-to-end encryption, and changes to technology models such as decentralised platforms'.[15]

7.6. We further note that the examples of the reasonable steps that could be taken by providers to meet the new expectations require further consideration. For example sub-section 8A(3)(c) says that a reasonable step is 'ensuring that training materials for generative artificial intelligence capabilities and models do not contain unlawful or harmful material' However, some services use AI models for internal use only (not made available for use by users of the service), for example to proactively find harmful content and services.

7.7. For these reasons DIGI considers that the inclusion of 8A is premature in the light of the collaborative work that will be undertaken by the Government and industry on a *voluntary AI Safety Standard* and additional consideration of the potential safety harms raised by emerging technologies in the context of the Statutory Review of the OSA. Arguably, regulating AI safety in subordinate regulation in advance of the review of the underlying legislation – in this case the OSA operates to constrain the independent reviewer's consideration of the safety of emerging technologies including the basic question of whether the OSA should remain technology neutral and whether that AI related harms are best dealt with under the auspices of the OSA or other regulatory frameworks. DIGI therefore recommends that the BOSE remains technology-neutral and outcomes-oriented and that the introduction of an expectation concerning AI is deferred pending the outcome of the OSA Review.

Recommendation

I. We suggest that proposed new expectation in sub-section 8A be deferred pending the outcome of the Statutory Review of the Online Safety Act 2021, which is scheduled to conclude in October 2024.

8. New expectation concerning recommender systems.

8.1. The amendments include two additional expectations in section 8B that providers of recommender systems:

1) *will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of recommender systems on the service.*

2) *will take reasonable steps to ensure that recommender systems are*

[15] *Terms of Reference – Statutory Review of the Online Safety Act 2021*, released February 2024.

designed to minimise the amplification of material or activity on the service that is unlawful or harmful.

8.2. We understand that the Government considers that recommender systems are '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.' We understand that the aim of the new expectations is to consider how the design and implementation of recommender systems on their service might impact end-users and incorporate safety measures to minimise any risk. The introduction of a new expectation concerning recommender systems raises the same concerns as the proposed new expectation on generative AI capabilities as it is a major change from the existing technology neutral policy settings of the OSA. The obligation on service providers to consider the safety impacts of recommender systems and mitigate risks are already covered by the core expectations in sub-sections 6(1) and 6(2) the BOSE.

8.3. We are concerned about the feasibility of the example reasonable steps concerning recommender systems. For example, a provided example of a reasonable step in sub-section 8B(3)(c) is 'enabling end-users to make complaints or enquiries about the role recommender systems may play in presenting material or activity on the service that is unlawful or harmful'. However, addressing individual complaints around recommender systems may be impractical for many services.

8.4. We note that the BOSE guidance already sets out guidance regarding the safety interventions that may be taken by providers of recommender systems as reasonable steps in accordance with subsections 6(1) and 6(2). Further, the OSA Review will, as noted above, consider this issue and we consider it inappropriate to make additional changes to the BOSE concerning recommender systems pending the conclusion of the review process.

Recommendation

J. We suggest that proposed new expectation in section 8B be deferred pending the outcome of the Statutory Review of the Online Safety Act 2021 which is scheduled to conclude in October 2024, noting that the BOSE guidance already addresses recommender systems in relation to the core expectation in sub-sections 6(1) and 6(2).

9. New expectation concerning transparency reporting.

A new expectation in sub-section 18A has been introduced that requires providers will *publish regular transparency reports, at regular intervals of no less than 1 month and no more than 12 months, with information regarding:*

a) *the service's enforcement of its terms of use, policies and procedures and standards of conduct mentioned in section 14;*

b) *the safety tools and processes deployed by the service (including in relation to a service's key features), and their effectiveness;*

c) *metrics on the prevalence of harms, reports and complaints, and the service's responsiveness; and*

d) *the number of active end-users of the service in Australia (including children) each month during the relevant reporting period.*

9.1. We understand that the purpose of this amendment is to 'provide valuable information to the Commissioner and the public in understanding what safety measures a service has in place and how effective those measures are, how services are enforcing their terms of use, policies and procedures and standards of conduct, and the prevalence of harms on a service'.[16] This expectation is problematic, not the least because the vast majority of providers of services that are within scope of this expectation will simply not be aware of or comply with these requirements because they are irrelevant to the nature and context of their services. Most designated internet services including websites and apps pose minimal or no safety risk to end-users. For example, it is difficult to see how a simple website that provides basic information about a business or an app in the form of a calculator, or a calendar or which provides weather information could meaningfully comply with these expectations. We consider that it is poor regulatory practice to draft regulatory requirements that are irrelevant to, and will be ignored by, the vast majority of services in scope.

9.2. Insofar as there is a need for the Commissioner to obtain access to this type of information, we note that the OSA contains ample data gathering mechanisms about the safety of services, including provisions that empower the Commissioner to request periodic reports and non-periodic reports concerning the BOSE under Part 4 of the OSA. Both the SMS Codes and the proposed RES/DIS standards also contain annual transparency reporting obligations to the eSafety Commissioner.

9.3. It is important that transparency is actually meaningful and provides data points and context that truly enable governments, civil society, and end-users to contextualise, understand, and advance online safety outcomes. Different audiences are likely to have different information needs. Transparency also needs to be balanced with competing priorities such as managing confidential business information and protecting safety systems and processes from exploitation by bad actors. We encourage thoughtfulness by the Department as the proposed amendments are considered, to ensure the BOSE powers can be effectively deployed to meet transparency needs, while maintaining the efficacy of safety measures and the confidentiality of business information.

9.4. Insofar as the public is concerned, we agree that it is reasonable to expect providers of services to make accessible relevant information to users about their safety tools and processes. We suggest that sub-section 18A be redrafted accordingly.

[16] *Amending the BOSE Determination Consultation Paper,* p.14

9.5. We note that the we consider there is a need to further consider, in the context of the Statutory Review of the OSA , whether eSafety' disparate powers to gather information and require information should be rationalised and administered in a more transparent and efficient manner, with clear and consistent metrics for 'like' categories of online harms.

Recommendation

K. The wording in subsection 18 be replaced by the following:

> Providers will make available to their users, relevant information about the safety tools and processes deployed by the service (including in relation to a service's key features).

10. Timeline of compliance with proposed amendments

We also wish to raise concerns about the timeframe for services to comply with any new expectations. The instrument outlines that any amendments, once registered, will commence the day after registration. We suggest that the government provide a reasonable timeframe of at least 6 months before the revised expectations are in force.

Recommendation

L. That the BOSE amending instrument provides a reasonable timeframe of at least 6 months before any amendments are in force.