To: Ms Delia Rickard PSM
By email: OSAReview@communications.gov.au

1 July 2024

Dear Ms. Rickard,

The Digital Industry Group Inc. (DIGI) appreciates the opportunity to provide you with our views on the statutory review of the Online Safety Act 2021 as advanced in the *Statutory Review of the Online Safety Act 2021 Issues paper*, April 2024 (*the Issues Paper*).

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

DIGI shares the Government's commitment to improving and promoting online safety in Australia. We acknowledge the powerful, pioneering work of the Office of the eSafety Commissioner in raising public awareness of the importance of online safety. Some of the greatest strengths of the Office of the eSafety Commissioner are:

- The development of and advocacy of the safety by design principles and framework.
- the highly efficient and cooperative arrangements that have been established and implemented with industry to ensure the prompt responses to takedown requests.
- the expert research conducted by the Office which has yielded nuanced perspectives on issues such as the exposure of young people to online pornography.
- the educational resources and tools provided to the Office which have provided guidance about online safety risks and empower users to manage them.
- the mechanisms that have been established to enable young people to consult with the Office on online safety issues that affect them.
- the work of the Office to promote dialogue and cooperation amongst overseas regulators.

Over the past 3 years, DIGI has extensively engaged with the Office, particularly in the development of the Phase 1 industry Codes (OSA Codes), six of which are now in force and require social media services, app distribution services, search engines, equipment providers and related services, hosting services and internet service providers to implement a range of binding systems and processes to address seriously harmful 'Class1' materials online (being categories of materials that would be refused Classification under the National Classification Scheme). The drafting of the OSA Codes was an intense effort involving more than 285 industry participants[1] over a 21-month period, co-led by DIGI with Communications Alliance, on behalf of the Australian online industry. DIGI has also engaged extensively with the Office on the development of the Standards for Designated Internet Services and Relevant Electronic Services (OSA Standards), and we are currently consulting with the Office concerning the development of Phase 2

[1] See *Request for Registration of Online Safety Codes* (revised 31 March 2023) available https://onlinesafety.org.au/codes/.

industry codes to regulate materials that are unsuitable for children and young people under 18 years old under the National Classification Scheme. These engagements have provided us with unique insights on the operation of OSA Codes which we have shared in this submission.

In this submission we have identified areas where we consider the OSA framework can be improved to better achieve its objectives. In particular, we have identified that there is scope to make significant structural improvements to the OSA to implement a more technology-neutral and scalable regulatory framework for online safety that is more appropriately targeted and better equipped to adapt to rapid technological changes and accommodate the different risk of harm on different services as well as the range of social harms to which Australians may be exposed on different services. We have also made a series of recommendations, as to how the OSA could be amended to achieve that aim.

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.

Yours sincerely,

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

Table of contents

1. Summary of recommendations **3**
Section 1: A principles based approach to future proof the OSA regulatory framework 4
Section 2: The objects of the OSA remain relevant and appropriate 4
Section 3: The OSA's approach to defining sections of the online industry is increasingly unworkable. 4
Section 4: The need to reduce regulatory overlap and complexity 5
Section 5 : The unsuitability of the National Classification scheme for regulating harmful materials online 5
Section 6: The need to strengthen the administration and governance of the BOSE. 5
Section 7: The process of drafting industry Codes 6
Section 8: The eSafety Commissioner's enforcement and Investigative powers 6
2. Discussion of Issues **6**
1. Adopt a principles based approach to future proof the OSA regulatory framework 6
Recommendations in Section 1 10
2. The objects of the OSA remain relevant and appropriate 11
Recommendations in Section 2 14

3. The OSA's approach to defining sections of the online industry is increasingly unworkable. 14
Recommendations in Section 3 17
4. The need to reduce regulatory overlap and complexity 17
Recommendations in Section 4 19
5. The unsuitability of the National Classification scheme for regulating harmful materials online 20
Recommendations in Section 5 24
6. The need to strengthen the administration and governance of the BOSE 25
Recommendation in Section 6 27
7. The process of drafting industry Codes 28
Recommendation in Section 7 30
8. The eSafety Commissioner's enforcement and Investigative powers 30
Recommendations in Section 8 30
3. Concluding Remarks 31
APPENDIX 32

1. Summary of recommendations

This section contains a summary of our recommendations for improving the legislation. The OSA review provides a valuable opportunity to reconsider fundamental questions about the scope of services and equipment regulated by the OSA, the increasing complexity of the scheme and its reliance on a range of overlapping subsidiary instruments as well as the specific questions raised in the *Issues Paper.* Our key recommendations are that that the reviewer give consideration to simplifying and strengthening the OSA by adopting:

- a risk based, proportionate and technologically neutral approach to regulation that can keep pace with rapid technological change; and
- enhanced mechanisms that improve the administration and governance of the Office's regulatory activities.

In making these recommendations, we note that currently the Government is conducting a range of concurrent regulatory processes that either directly influence or intersect with this review and have the potential to substantially influence the services in scope of the OSA, and the material subject to the provisions of the OSA. This context poses challenges for stakeholders to provide detailed feedback to the extent that these interrelated concurrent processes have only recently or not yet concluded. We recommend that the Department can provide information about the coordination and sequencing of the inter-related processes which are set out in the Appendix to this submission.

Section 1: A principles-based approach to future proof the OSA regulatory framework

A. We recommend that the OSA adopt a proportionate, risk based, approach to improving safety online for Australians that is based on a common set of mandatory accountability requirements for Australian online services and equipment providers that replace the current regime of subsidiary regulatory instruments i.e. BOSE, Codes and Standards.

B. In addition, we recommend that the scope of the OSA scheme be revised to reflect the above proportionate risk-based approach. This should include removing from scope services which have limited functionality or are low risk.

C. We also recommend the OSA be amended to require that the eSafety Commissioner have regard to the need for regulatory activities to:

   a. be proportionate to the risk of harm online to end-users in Australia.

   b. take into account the technical capability of particular providers of services or equipment to reduce the risk of harm online to end-users in Australia.

   c. the importance of protecting and promoting human rights online, including the right to freedom of expression, the right not to be subject to arbitrary or unlawful interference with privacy, the right to protection from exploitation, violence, and abuse, including associated statutory obligations.

   d. where appropriate, be informed by consultation with government, regulatory, commercial, consumer and other relevant bodies and organisations; and

   e. be transparent, targeted, and consistent.

Section 2: The objects of the OSA remain relevant and appropriate.

D. We recommend that the objects in section 3 of the OSA should be retained.

E. The regulatory requirements OSA should, as far as reasonably practicable, be technologically neutral to future proof the regulatory framework against rapid technological change.

Section 3: The OSA's approach to defining sections of the online industry is increasingly unworkable.

F. As outlined in Section 1, we recommend that the OSA scheme be revised to remove from scope services with limited functionality or which are low risk to align with a risk-based approach.

G. The current categories of services and equipment in scope of the OSA are not workable and should be replaced by an approach that takes into account variances in their risk profiles and legal and technical capabilities.

Section 4: The need to reduce regulatory overlap and complexity.

H. To the extent that the OSA framework continues to regulate via the BOSE, Codes and Standards we recommend that the OSA be amended to clarify that:

   a. The Codes and Standards are the appropriate regulatory scheme for minimum mandatory obligations aimed at high-impact content.

   b. The BOSE instrument is a flexible set of expectations that providers are required to report against.

Section 5: The unsuitability of the National Classification scheme for regulating harmful materials online

I. Rather than regulating materials in accordance with the National Classification Scheme, we recommend that the OSA regulate materials that are lawful but harmful in relation to a standard of 'harmfulness,' rather than the current standard of 'offensiveness.'

J. We recommend that the OSA contain clear criteria for categories for harmful materials such as pornographic material that require child specific regulation, such as age restriction.

K. Rather than regulating harmful materials via a range of overlapping and conflicting subsidiary instruments, we recommend that consideration be given to introducing an enforceable, proportionate standard that services must implement and enforce terms of use that prohibit harmful content, having regard to their risk profile.

L. If the OSA is to continue to rely on the criteria for classifying materials under the National Classification Scheme, it should be amended to:

   a. require the Commissioner to publish clear guidelines for the range of content that falls to be Classified under the OSA, with input from the National Classification Board and in consultation with academic experts, human rights and civil society representatives and the broader community.

   b. introduce a mechanism to record decisions by the Commissioner to classify material; as Class1 or Class 2, which can be accessed by industry participants and the public.

   c. require independent oversight of the Commissioner's decisions, by an independent board, which represents community interests including establishing an independent mechanism via e persons impacted by decisions of the Commissioner to classify material as Class1 or Class 2 to seek a review of the decisions.

Section 6: The need to strengthen the administration and governance of the BOSE.

M. The administration of the BOSE should be improved by:

   a. Including a mechanism in the OSA via which providers who are the subject of BOSE notices can seek protection for commercial in confidence information.

b. Introducing a requirement in the OSA for the eSafety Commissioner to implement and publish an annual plan that sets out its annual program of regulatory activities.

c. Introducing an express right for a provider who has responded to a BOSE notice to be given a draft of the eSafety Commissioner's intended report prior to publication/public comment and an ability to challenge findings before they are made public.

d. Limiting the scope of BOSE notices to requiring providers to provide an explanation of the steps they are taking to meet relevant measures rather than giving them a limited opportunity to confirm whether or not they are taking a specific step that is not expressly mentioned anywhere in the BOSE itself.

e. Establishing a BOSE advisory Board with representation from industry, civil society and experts concerning the annual plan in b.

Section 7: The process of drafting industry Codes

N. We recommend that the Class 2 Code development process be deferred until the OSA review concludes and any relevant recommendations have been considered by the Government. Alternatively, we ask the reviewer to consider recommending that at least 12 months is allowed for these Class 2 Codes to be developed.

Section 8: The eSafety Commissioner's enforcement and Investigative powers

O. We recommend that any proposed changes to the Commissioner's regulatory and investigatory powers should be based on a thorough evaluation of effectiveness of the current enforcement and penalty regime and should be primarily directed at addressing identified deficiencies with the effectiveness of the current regime.

2. Discussion of Issues

This section contains DIGI's feedback concerning the operation of the current OSA regulatory framework, including key questions raised by the Issues Paper.

1. Adopt a principles-based approach to future proof the OSA regulatory framework

1.1. The OSA was updated in 2021 with a vision to "create a modern, fit for purpose regulatory framework that builds on the strengths of the existing legislative scheme for online safety".[2] In DIGI's view, the 2024 OSA review is an opportune moment to reconsider the foundational purpose and objectives of the OSA given the increasing complexity of the regulatory framework since it was introduced. The challenges in

[2] *Online Safety Act 2021 Explanatory Memorandum.*

developing online safety regulation that is fit for purpose are well articulated by the Digital Trust & Safety Partnership:

> *there is no one-size-fits-all approach to handling online content and associated behavioural risks or, more generally, to companies' Trust & Safety operations. Depending on the nature of the digital service, each may face unique risks relative to the various products or features they provide — different threats, different vulnerabilities, and different consequences. Products or features may engage with end users directly or indirectly, as well as with other services or businesses. What is an effective practice for one digital service may not suit another, and highly prescriptive or rigid approaches to defining Trust & Safety practices are likely to be too broad, too narrow or have negative unintended consequences. Further, risks change over time and so approaches to mitigating them must also have room to evolve*[3].

1.2. Any proposal to expand or improve the OSA framework should start from a thorough assessment of the operation and effectiveness of the existing regulatory scheme. This requires consideration of how best to measure the efficacy of Australia's online safety regulatory framework as it is today. We consider that the success of the OSA framework should be evaluated by measuring the extent to which it can prevent Australians from being harmed online. This requires a clear understanding of what the concept of harm captures. This is at present difficult to measure because while the OSA contains specific requirements on providers to address the risk of harm (for example, see core expectations in sub-section 6(1) of the BOSE), the concept of 'harm' is not clearly defined in the OSA. While harm is undefined and harm prevention is very difficult to measure, we suggest that a helpful starting point for the reviewer may be to consider what is the appropriate scope of harm under the OSA and what are the key components of successful harm prevention and if these elements are present in the current regulatory approach. For example, we would suggest that the concept of harm in the OSA be focused on psychological and physical harms to users rather than economic harms, societal harms or reputational harms that are dealt with by other legislative frameworks. Further we suggest that the framework should incentivise providers to evaluate the risk of harm to users and adapt their response to harm in a manner that is proportionate to that risk.

1.3. We also consider that it is important that the reviewer consider the extent to which key elements of the scheme are brand new, or not yet in place[4]. As outlined in the remainder of this submission, the current approach to regulation of online safety under the OSA attempts to catch all services and equipment used by Australians to access the internet, while prescribing in subsidiary legislative instruments very detailed regulatory requirements for specific categories and subcategories of services and equipment and specific types of technologies. This has resulted in a complex framework with numerous overlapping requirements. We ask that the reviewer take this context into account.

Question 7 of the *Issues Paper* asks: Should regulatory obligations depend on a service provider's risk or reach?

[3] Digital Trust and safety partnership, *Trust & Safety Best Practices Framework* 2021 accessed at https://dtspartnership.org/best-practices/.
[4] For example, Class 2 Codes and the periodic notices for BOSE are not yet in place.

1.4. DIGI's principal recommendation in this submission is that the reviewer give consideration to restructuring and narrowing the OSA scheme, **based on a proportionate, risk-based based and technologically neutral approach to regulation that is focused on protecting Australians from online harms, consistent with the current objects of the Act**.

1.5. Specifically, we recommend a review of in scope services and the introduction of a proportionate, risk-based approach to improving safety online for Australians that is based on a common set of accountability requirements for Australian online services and equipment providers. We think that the reach of a service (both within Australia and globally) is a relevant consideration of the risks posed by certain service categories such as social media services.

1.6. However, the reach of services is not an overriding consideration in assessing risk and factors such as the purpose and functionalities of a service should also be taken into consideration. For example, a common mobile app is a calculator, but even where that app is widely distributed it should be considered as presenting no safety risk and therefore out of scope because of its core purpose and absence of interactive functionality. The reviewer should reconsider other low risk services such as email, which is omitted from scope of both the EU's DSA and UK's Online Safety Act for this reason.

Question 31 of the Issues Paper asks what features of the Act should be expanded? Question 22 asks Should Australia place additional statutory duties on online services to make online services safer and minimise online harms?

1.7. DIGI recommends that the reviewer consider how the overall effectiveness of the legislative framework can be strengthened by adopting a less complex, more streamlined regulatory approach that replaces the current scheme of overlapping subsidiary legislative instruments. We consider that an effective risk-based approach could be founded on a streamlined and common set of enforceable standards for in-scope services with similar functionalities, focused on requiring regulated entities to take proportionate steps to minimise online safety risks based on their risk profile. These standards could draw upon some of the common principles that underlie the current suite of subsidiary regulatory instruments (BOSE, OSA Codes and OSA Standards). We suggest that this should be complemented by some basic principles that guide the exercise of the eSafety Commissioner's discretion under the OSA as discussed in 1.8. The advantage of this approach is that it:

1.7.1. reduces the need for prescriptive, technology-specific regulations of different categories and subcategories of online services, which must be frequently updated to remain fit for purpose.

1.7.2. overcomes the unacceptable level of fragmentation, complexity, and uncertainty within the current regulatory framework.

1.7.3. has the advantage of aligning with similar international regimes such as the European Union's Digital Services Act (DSA).

1.7.4. enables industry to respond to the results of expert research e.g. about the effectiveness of industry measures and evolving threats and trends.

1.7.5. provides greater certainty to industry and users regarding the way in which the Commissioner's discretionary powers will be exercised.

1.7.6. Is more harms agnostic and able to respond to new and emerging challenges without needing such frequent reviews of the primary and secondary regulations which are resource intensive and time consuming for government.

1.7.7. is more readily enforceable than a broad-based duty of care given the inherent challenges of enforcing that duty by demonstrating that a lack of care has resulted in objectively ascertainable injury to users. When the courts determine whether a duty of care has been breached, they take into account proximity of relationship between the persons by whom and to whom the duty is said to be owed. In this respect, we note that analogy in the *Issues Paper* to an online duty of care and Work Health and Safety law is unhelpful since the employer's duty of care arises from the fiduciary relationship between employer and employee. This special trusted relationship does not correspond to the more distant relationship between users and online service providers. Furthermore, many of the types of harms to which users can be exposed online are of a different nature and are often the result of users' choices and behaviours towards each other.

Question 29 of the Issues Paper asks: is the current level of transparency around decision-making by industry and the Commissioner appropriate? If not, what improvements are needed?

1.8. In addition to the overly broad reach of the regulatory framework, DIGI is concerned that the OSA contains very limited mechanisms to promote good regulatory governance and practice by the Commissioner in carrying out regulatory activities under the OSA. In particular, the OSA does not contain sufficient transparency and accountability requirements concerning the exercise of the Commissioner's discretionary powers. While Section 183 of the OSA imposes some minimum reporting obligations on how often the Commissioner is utilising powers under the Act, industry participants do not have sufficient clarity about how their compliance with regulatory requirements is assessed. There is, for example, no requirement on the Office to publish guidance about how it approached the classification of Class 1 and Class 2 online materials. Furthermore, the OSA provides very limited opportunities and strong disincentives for industry participants to challenge the Commissioner's decisions, which may encourage some participants to take an over-cautious approach to removal of content.

1.9. DIGI also notes that the Office's approach to consultation has varied widely. There have been notable instances of strong consultation and engagement between industry, key stakeholders and the Commissioner that has resulted in very positive outcomes. For example, the Office's consultation process for the development of the safety by design principles[5]. Similarly, it is clear that the Office took a very considered approach to revising the OSA standards following the public process which they have helpfully documented. There have also been notable instances where in DIGI's view the Office has not

[5] Office of the eSafety Commissioner, *Safety by Design Overview* (May 219).

consulted, where this would have been expected and helpful to ensure confidence in the Commissioner's decision-making process. For example, the current BOSE Guidance was issued by the Office without consultation with industry, although when the BOSE was first released the Department advised industry that this would occur.[6] Section 7(2) of the BOSE states that providers must have regard to any relevant guidance material published by eSafety. Given the status of the guidance, we hope that industry is afforded an opportunity to provide input into future versions of the guidance. This has a considerable impact on providers given the highly prescriptive nature of the guidance which is often framed in mandatory terms (see discussion in section 4 below). Similarly, we have recommended stakeholders be afforded an opportunity to consult on updates to the Position Paper for Code development to facilitate the Phase 2 Code development process.

1.10. In the remainder of this submission, we have identified additional areas where the administration and governance of the OSA can be improved to promote regulatory best practice by the Office of the eSafety Commissioner. To that end, we recommend that the OSA should set out the key considerations which should inform its regulatory activities under the Act.

[6] Department of Infrastructure, Transport, Regional Development and Communications, *Frequently Asked Questions Basic Online Safety Expectations October 2021,* p.2.

Recommendations in Section 1

B. We recommend that the OSA adopt a proportionate, risk based, approach to improving safety online for Australians that is based on a common set of mandatory accountability requirements for Australian online services and equipment providers that replace the current regime of subsidiary regulatory instruments i.e. BOSE, Codes and Standards.

C. In addition, we recommend that the scope of the OSA scheme be revised to reflect the risk-based approach. This should include removing from scope services which have limited functionality or are low risk.

D. We also recommend the OSA be amended to require that the eSafety Commissioner have regard to the need for regulatory activities to:

   a. be proportionate to the risk of harm online to end-users in Australia.

   b. take into account the technical capability of particular providers of services or equipment to reduce the risk of harm online to end-users in Australia.

   c. the importance of protecting and promoting human rights online, including the right to freedom of expression, the right not to be subject to arbitrary or unlawful interference with privacy, the right to protection from exploitation, violence, and abuse, including associated statutory obligations.

   d. where appropriate, be informed by consultation with government, regulatory, commercial, consumer and other relevant bodies and organisations.

   e. be transparent, targeted, and consistent.

   f. in relation to children and their interests, be informed by the best interests of the child.

2. The objects of the OSA remain relevant and appropriate.

Question 1 of the *Issues Paper* asks: Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?

2.1. DIGI supports the retention of the objects in section 3 of the OSA to improve and promote online safety for Australians. Section 3 is broadly framed and enables the Office of the eSafety Commissioner to engage in a wide range of regulatory, enforcement and educational and advocacy initiatives, both within Australia and internationally.

Question 27 of the Issues Paper asks: Should the Commissioner have powers to act against content targeting groups as well as individuals? What type of content would be regulated and how would this interact with the adult cyber-abuse and cyberbullying schemes?

2.2. DIGI considers that the OSA should focus on the safety of individuals, rather than wider societal harms such as mis and disinformation[7] and hate speech that targets groups of people (rather than individuals) on the basis of protected characteristics.[8] DIGI considers that this focus is appropriate and consistent with international approaches to online safety regulation. We support the current approach to regulating online abuse including online hate that targets an individual under the cyber-bullying scheme. We recommend that the Government's regulation of societal based harms such as dis- and misinformation and hate speech targeting groups, should be dealt with in other legislative frameworks which can better target the services where this content occurs and these areas often warrant a more nuanced approach that balances freedom of speech considerations than that which is provided by the OSA[9].

2.3. There is currently considerable uncertainty about the scope of harms captured by the OSA and the standard which industry must meet to comply with regulatory requirements. To take a few examples that reflect the uncertainty that arises from the lack of definition at an OSA level, the *Basic Online Safety Expectations Regulatory Guidance* issued by the eSafety Commissioner in September 2023 (BOSE Guidance):

2.3.1. indicates that the requirement in expectation 6(2), for instance, is "broader" than the defined categories of material covered by the OSA.

2.3.2. references harm that occur at a societal level, as well as at an individual level (see page 30 for example); and

2.3.3. when discussing expectation 14, effectively require providers to remove all "activity and material that is unlawful and harmful" by having and enforcing a prohibition of this nature via their terms of use, noting that this should include the specific harms set out in expectation 13, but that providers should also consider other harms and "should update their terms of use, standards of conduct and other policies and procedures as new risks and harms emerge over time".

2.4. Of course, many providers do in fact have prohibitions on unlawful and harmful activity and material in their terms of use or acceptable use policies and take action in response to a range of harms. These are a key tool for providers to promote online safety. We also acknowledge that the BOSE and the BOSE Guidance do anticipate that enforcement action by a provider may vary depending on the nature of a user's breach. However, exposing a provider to possible regulatory action for a failure to enforce a prohibition on all forms of unlawful and harmful activity and material effectively delegates decisions regarding what harms *should* be prohibited and properly within the scope of the OSA and the remit of the Office of the eSafety Commissioner without the

[7] Currently the subject of the proposed *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill* (2023)

[8] Hate speech can be an individual harm to the extent it targets particular individuals and a collective harm e.g. hate speech that is targeted at harming a group or Class of people. Protected characteristics may include race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease.

[9] See DIGI's submission on exposure of Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023, August 2023 and DIGI Submission on revisions to updated Basic Online Safety Expectations 2023 dated 26 February 2024.

parliamentary debate that would normally surround such a decision. It also places providers in the extremely difficult position of having to make such a decision, often in a context where the topic is contested within the community, placing providers under threat of regulatory action for a perceived failure to treat a particular harm in the way that the eSafety Commissioner views as appropriate to meet expectations.

2.5. DIGI recommends that the set of harms within the scope of the OSA, and eSafety's remit, should be clearly set and defined by parliament. This will limit uncertainty within the OSA scheme itself, but will also avoid conflict and overlap with other regulatory regimes and regulators given there are obviously a broad range of online harms that are properly the subject of other regulatory schemes (e.g. privacy harms regulated by the OAIC under the Privacy Act, or consumer harms regulated by the ACCC under the Competition and Consumer Act). The potential for overlap was evident last year when the exposure draft *Communications Legislation Amendment (Combatting Misinformation and Disinformation Bill 2023)* also purported to address a number of online harms, whilst addressing potential overlap by stating (at section 36) that a misinformation code or misinformation standard would have no effect to the extent that it dealt with a matter dealt with by the BOSE (which itself, as outlined above, is subject to considerable uncertainty as to scope given the undefined references to online harms).

Question 29 of the Issues Paper asks Should the Act address risks raised by specific technologies or remain technology neutral?

2.6. We think that it is important that the OSA adopts a technologically neutral approach to ensure that it remains fit for purpose. We consider that the introduction of prescriptive requirements for evolving new technologies is unsustainable given the pace of technological change. In addition to retaining the focus on individual harms, DIGI considers that the OSA should remain technologically neutral. However, the OSA's scope should be reviewed in line with recommendations in section 3 below.

Question 3 of the Issues paper asks: does the Act regulate things (such as tools or services) that do not need to be regulated, or fail to regulate things that should be regulated?

2.7. A range of new requirements for new sub-categories of AI enabled services to have already been introduced into the BOSE, Phase 1 Code for search engine services and Phase 1 industry standards for DIS. For example, the Phase 1 Search Engine Code services code contains requirements related to AI features of those services. The OSA Standards further specific requirements for sub-categories for 'High impact generative AI DIS' and 'machine learning model platforms. Similarly, under the BOSE, the Minister has created new expectations for sub-categories of services: 'service that use or enables the use of generative artificial intelligence" capabilities and services that "use recommender systems"[10]. In addition, the Issues paper proposes that a range of additional technologies could be addressed by amendments to the OSA including:

[10] Expectations 8A and 8B of the Basic Online Safety Expectations

- generative artificial intelligence.
- immersive technologies.
- recommender systems.
- end-to-end encryption; and
- changes to technology models such as decentralised platforms[11].

2.8. While we understand that online harms can and do emerge on these services, the usage of the services are broad and approaches to the harms will be best addressed through a technologically neutral approach. It is also important to appreciate that while new technological developments may entail risks, there are also opportunities to use new technologies such as AI to help improve online safety[12]. This includes opportunities to improve the safety of online services for Australians. For example, while generative AI can be used to perpetuate harmful deep fake imagery online, this technology can be used to streamline platforms' efforts to detect and disrupt harmful materials. For example, Microsoft is integrating GPT4 enabled content moderation solutions for Microsoft Start service to proactively block content that violates content policies[13]. It is important that policy settings do not discourage these types of innovations. By keeping the focus on harms rather than on specific technologies the OSA will be able to adapt to both the risks and opportunities presented by new technological developments without the need for multiple legislative amendments to keep pace with new technological developments.

Recommendations in Section 2

E. The objects in section 3 of the OSA should be retained.

F. As recommended in response to Section 1above, we recommend that the scope of the OSA scheme be revised to remove from scope services with limited functionality, or which are low risk, to align with a risk-based approach.

3. The OSA's approach to defining sections of the online industry is increasingly unworkable.

Question 2 asks, does the Act capture, and define the right sections of the online industry?

[11] Issues Paper p 51,52.
[12] Department of Science, Industry and Resources, *Safe and responsible AI in Australia consultation Australian Government's interim response*, 2024 p.4.
[13] See Microsoft ACPDM Transparency report 2024 available at https://digi.org.au/disinformation-code/transparency/

3.1. DIGI considers that the current approach of the OSA to regulating categories of the online industry is increasingly unworkable and should be replaced by **a proportionate, risk-based, and technology-neutral approach to regulation as outlined in section 1.**

3.2. In contrast to similar regulatory frameworks in the UK and the EU, the OSA adopts a "catch-all approach," providing the eSafety Commissioner with regulatory powers over all the points at which consumers access the internet regardless of the level of risk or even the presence of risk at all. This includes ISPs, search engines, social media, hosting services, email and messaging services, gaming services, dating services and apps and websites as well as providers of equipment and equipment related services. As currently structured, the OSA defines eight industry sections that are subject to different sets of regulatory requirements with the designated internet services category regulating any service not captured by the other industry sections. There are key differences between categories of services and equipment in scope of the OSA (eg public facing service, and private communications services) and significant policy surrounding differences in how those services are regulated - grouping all services in scope in an equivalent way at a statutory level without any decisions being taken by parliament as to how such differences should be reflected in the regulatory regime effectively delegates decisions on these types of matters to the eSafety Commissioner.

3.3. In practice the current approach is increasingly unworkable, because:

3.3.1. Regulation that is focused on the category of service makes it very difficult to ensure that the OSA captures the 'right services' based on the risk of harm to Australian end-users. Many services regulated by the OSA do not pose a risk to users that warrants regulation. By far the majority of websites and apps with which users interact are safe and pose little or no risk of harm to users. Similarly, the Government recognises that many applications of AI do not present risks that require a regulatory response[14]. Whilst some work has been done to introduce a more risk-based approach into Codes and Standards, and some but not all of the BOSE expectations require "reasonable steps" to be taken on a particular point, which *should* operate to enable providers to apply measures in a way reasonably adapted to the level of risk, this is not embedded at a statutory level and in practice all providers in Australia are subject to the regime. The current approach relies on the regulator's discretion not to enforce requirements of the OSA against the majority of services in scope, which creates a high degree of uncertainty for industry participants.

3.3.2. As defined in the OSA, the boundaries between some service categories are inherently unclear, for example, it is currently difficult to differentiate between social media services and relevant electronic services (RES) to the extent that some services such as dating services or messaging services may involve elements of communication and social interaction. Some service category definitions in the OSA have significant potential for overlap (for example, the definition of a designated internet service (DIS) excludes some but not all other service categories) and some (e.g. internet search engine services) are missing altogether.

3.3.3. The categories of services and equipment that are in scope of the OSA are and are likely to remain in a constant state of evolution, with new services emerging

[14] *Safe and responsible AI in Australia consultation Australian Government's interim response* p.5.

in response to new technologies. As a result, the boundaries between the current categories of services are increasingly unclear. For example, with the advent of generative AI, many services will continue to incorporate new features in their services in order to enhance their users' experience in different ways, and for different purposes (personal, commercial, educational). However, it is unclear the extent to which integration of generative AI functionality in a service changes the category of a service for example from a search engine service to a designated internet service.

3.3.4. Some categories of services under the OSA are so broad and bundle together services that are very diverse, and have very different functionalities, levels of interactivity and levels of safety concern. This has made it very challenging to meaningfully regulate for the category as a whole. In particular, we note the diversity of services that fall within the relevant electronic services (RES) and designated internet services (DIS), illustrated by the large number of sub-categories that are contained within the Industry Standards, all with different measures being applicable. The RES category includes services as diverse as email services, MMS, and SMS services, dating services, gaming services with communications functionality, and messaging services with varying reach, and risk profiles. DIS services are even more diverse again, and include operating systems, cloud storage, and the full gamut of apps and websites used by Australians ranging from simple apps that help users find the date or time of day, sites that provide community services to pornography sites.

3.3.5. The breadth of the DIS and RES categories of services and the equipment category regulated by the OSA has led to increasing fragmentation of the categories of the scheme in the scheme's subsidiary instruments. For example, the OSA Codes and OSA Standards for RES and DIS introduce a range of new sub-categories of services and equipment so as in order to ensure the regulatory requirements for those instruments are commensurate with the risks and technical and legal capabilities of the diverse service and equipment types in scope. While the industry did its best to work with the OSA scheme to develop codes for RES and DIS, these codes were ultimately not registered.

3.3.6. The technical and legal capability of different service providers to identify, assess and respond to different types of harmful materials and activity also vary widely within the current categories and subcategories of regulated services. While the OSA, to a certain extent, acknowledges that differences in technical capability may impact on the ability of certain service categories to respond to a takedown request under the Online Content scheme, there is a need for the different capabilities of services to be factored into the regulatory framework more generally.

3.4. We note that the varying technical capability of different services was identified by DIGI and Communications Alliance as a key issue, and an industry learning, in the development of the Phase 1 OSA industry Codes. For example, DIGI has identified that:

3.4.1. Enterprise-based services such as those provided to the Government, will of their nature be low risk but also subject to stringent contractual arrangements between service providers and their customers that protect the security and integrity of the enterprises data that limit the extent to which the service provider

can intervene in relation to the activities of the enterprise customer. Enterprise customers themselves are customers of technology providers, not technology companies themselves, and cannot be treated as technology providers.

3.4.2. The majority of apps and websites accessed by Australian end-users are low risk for Class 1A and Class 1B materials , including retail websites, websites containing contact and service information for small businesses such as cafes, hairdressers and plumbers, apps offered by medical providers to allow patients to access x-ray imagery, information apps such as weather, time and calculator apps, train or bus timetable apps, newspaper websites, personal blogs, and artistic websites.

3.4.3. Due to the nature of relevant electronic services (RES), their functionality, and the manner in which they are otherwise regulated, not all service providers within these category services will be capable of reviewing and assessing material communicated by end-users on their services. Further, as these services often provide an ability for end-users to communicate with end-users of other services, providers will not always have a contractual relationship with all end-users involved in a communication, or visibility surrounding any engagement between an end-user involved in a communication and another service provider. ISPs are legally prohibited from interrupting communications between users of SMS/MMS and email services they provide, which also impacts on their ability to identify, assess, and remove Class 1 materials under the OSA[15]. The capability of different providers to comply with regulatory requirements that require the provider to access and view individual items of material, and end-user activity, will vary widely depending on their legal and technical capacity to do so.

3.4.4. End to end encrypted services cannot scan users' communications for child sexual abuse imagery, for example by using Photo DNA technology, without compromising the security of data communications on the service afforded by end-to end-encryption[16]. Encryption is a vital part of modern electronic communications as it allows two or more parties to securely and confidentially engage with each other in many forms of communication and online activities. The ability to encrypt (and subsequently decrypt) communications underpins almost every online activity, from speaking on a mobile phone, accessing Government services to online banking, shopping and web browsing. It is fair to say that most of the common online activities that so many Australians engage with numerous times each day would not exist in their current form, or not at all, if not for the security that encryption affords.

3.4.5. App distribution service providers do not directly control or have visibility of all content shared via third-party apps distributed via the provider's app distribution

[15] *Telecommunications (Interception and Access) Act 1979.*

[16] Note that the Government currently has significant powers of surveillance to detect unlawful activities online. For example, the Assistance and Access Act 2018 (Cth), International Production Orders Act 2021 (Cth) and the Surveillance Legislation Amendment (Identify and Disrupt) Act 2021(Cth) provide law enforcement agencies and intelligence organisations with far-reaching powers to access any network, system, device, or user accounts covertly and, where required, with the assistance of the service provider.

service and cannot take direct action to prevent access or exposure by Australian end-users to Class 1 materials via such apps.

3.4.6. Downstream search engine services do not have legal or operational control of the search functionality - in particular, the index and results served to users - on the search engine technology they licence.

3.5. The differences in capabilities for different services do not mean providers cannot take effective steps to identify risk and take mitigating action to improve the safety of their service for Australian end-users. However, certain types of intervention will not be appropriate for all services. For example, some encrypted services may be able to deploy a range of measures that disrupt and deter users from sharing child sexual abuse imagery. However, encrypted service providers need sufficient flexibility to tailor those measures to work with the different architecture of encrypted services.

Recommendations in Section 3

G. The current categories of services and equipment in scope of the OSA are not workable and should be replaced by an approach that takes into account variances in their risk profiles and legal and technical capabilities.

4. The need to reduce regulatory overlap and complexity.

Question 3 asks, does the Act regulate things (such as tools or services) that do not need to be regulated, or fail to regulate things that should be regulated?

4.1. DIGI notes that the regulatory framework of the OSA has become increasingly complex, and difficult for industry participants to navigate due to the layering of overlapping requirements in different subsidiary regulatory instruments and guidance materials that cover the same subject-matter. This overlap has created a range of potential conflicts in the regulatory framework which are difficult for industry participants to reconcile and create considerable uncertainty as to the standard of compliance expected of industry participants.

4.2. In particular, the recently amended BOSE includes a range of new provisions that overlap with, but are somewhat inconsistent with, requirements of the SMS Code and the OSA Standards. 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. 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:

4.2.1. Generative AI.

4.2.2. Risk assessments.

4.2.3. Complaint mechanisms for end-users.

4.2.4. Investing in systems, tools, and processes to improve the prevention and detection of material or activity on the service.

4.2.5. Proactive detection of materials.

4.2.6. Enforcement of terms, policies, and/or procedures.

4.2.7. Transparency reporting.

4.3. 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.*[17]

4.4. Furthermore Part 4 of the BOSE Guidance is framed in terms of things that a provider "should" or "should not" do, such that a provider risks being viewed as failing to meet expectations if it does not comply with the examples outlined in the BOSE Guidance. The BOSE are increasingly being treated in practice by the Office of the eSafety Commissioner as specific and mandatory or quasi-mandatory obligations rather than flexible principles that underpin transparency reporting. The BOSE Guidance, therefore, is difficult to reconcile with the original regulatory intention for the BOSE and Code as set out in the *Online Safety (Basic Online Safety Expectations) Determination 2021 Consultation Paper* published in July 2021 (BOSE Consultation paper).

4.5. We understand that the BOSE were intended to be flexible statutory expectations underpinned by transparency reporting. The BOSE Consultation Paper explained that the BOSE 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 OSA Codes and OSA Standards were 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.*

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

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.[18]

4.6. This type of overlap and conflict in the OSA framework is likely to continue to the extent that the OSA contains two separates, 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). In some instances, inconsistency arises even within a single regulatory instrument. To take one example only, there is inconsistency in the way class 1B material is treated in the new OSA Standards. Some provisions require a provider to effectively prohibit both class 1A and class 1B material on services, whereas others require the same services to prohibit class 1A material whilst class 1B material may be either prohibited, or subject to restrictions, on the service (reflecting the fact that not all class 1B material is prohibited offline in the same way as CSAM and pro-terror material).In section 1 of this part of the submission, we have recommended that the OSA adopt a more streamlined, principled, risk based approach that replaces these instruments. However, if the BOSE, Codes and Standards are retained, we recommend that the different function and role of each type of instrument under the OSA be carefully defined in the OSA and maintained to clearly demarcate their different purposes.

Recommendations in Section 4

H. To the extent that the OSA framework continues to regulate via the BOSE, Codes and Standards we recommend that the Act be amended to clarify that:

   a. The Codes and Standards are the appropriate regulatory scheme for minimum mandatory obligations aimed at high-impact content.

   b. The BOSE is a flexible set of expectations that providers are required to report against.

5. The unsuitability of the National Classification scheme for regulating harmful materials online

5.1. The OSA currently confers a range of powers on the eSafety Commissioner with respect to regulate 'Class 1" and "Class 2" materials. These include powers that enable the Commissioner to require services to remove Class 1 online material and to restrict/remove Class 2 material (which under the National Classification Scheme is unsuitable for under 18-year-olds). The BOSE also contain expectations that relate to Class 1 and Class 2 materials, including expectations that providers take reasonable steps to minimise the provision of Class 1 materials (expectation 11) and to ensure that technological or other measures are in effect prevent access by children to Class 2 material provided on the service (expectation 12). In addition, the eSafety Commissioner has powers to request the development of and enforce industry codes to regulate Class 1 and

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

Class 2 online materials, failing which the Commissioner can develop and enforce industry standards.

5.2. By way of background, offline content is subject to the National Classification Scheme which is a cooperative arrangement between the Australian Government and state and territory governments for the Classification of films, publications, and computer games. The National Classification Code and the guidelines for the Classification of films, computer games and publications were designed primarily for the assessment of commercially produced material before its release into the community.[19] Under the National Classification Scheme, the content is largely classified having regard to its 'offensiveness'.[20] The National Classification Code, guidelines for the Classification of films, computer games and publications provide the principles and criteria for making Classification decisions.[21] Under the OSA , Class 1 and Class 2 material is defined by reference to:

- the Classification it has received by the Classification Board under the Classification Act (where the material has been Classified), or
- eSafety's assessment of "the Classification the material would likely be given by the Classification Board under the Classification Act (where the material has not been Classified)."[22]
- The eSafety Commissioner is also able to request advice from the Classification Board on whether particular material is Class 1 or Class 2, but the Commissioner is not required to submit content to the Board for review.[23]

5.3. Of ongoing concern to the broad scope of businesses currently regulated by the OSA is the extent to which the regulation of online content under the OSA is tied to the National Classification Scheme, including the Classification Guidelines (the Guidelines). As outlined in the *Public Consultation Paper: Modernising Australia's Classification Scheme - Stage 2 Reforms*, April 2024:

> *Definitions of content to be Classified under the Classification Act, however, were developed in a predominantly physical media environment. Since this time, the rapid growth in online content and the emergence of new digital platforms for distributing content have posed challenges for the Scheme.*[24]

In particular, the criteria for the classification of content under the Guidelines, were developed for the regulation of specific categories of professionally produced material before its commercial release (films, computer games and publications), rather than for the regulation at scale of the infinite every day online personal, business and governmental interactions by online intermediaries including private communications that occur on services as diverse as email and

[19] eSafety Commissioner, *Development of industry codes under the Online Safety Act, Position Paper*, September 2021 p. 18.
[20] *Ibid* pp 20/21,
[21] Refer to https://www.Classification.gov.au/about-us/legislation as accessed on 18 Nov 2022.
[22] eSafety Commissioner, *Development of industry codes under the Online Safety Act, Position Paper*, September 2021 p. 19,
[23] Section 160, OSA.
[24] Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *Public Consultation Paper: Modernising Australia's Classification Scheme - Stage 2 Reforms*, April 2024 p. 8.

messaging services, hosting services, apps, social media services, streaming services and websites.

5.4. Further, the basic philosophy of the OSA is to regulate online harms whereas the basic philosophy of the Scheme is grounded in guiding principles that: "adults should be able to read, hear, see and play what they want; children should be protected from material likely to harm or disturb them; and everyone should be protected from exposure to unsolicited material that they find offensive."[25] Consequently, the key criteria for assessing content under the Classification Scheme, a Classification Board is the offensiveness of material, rather than its harmfulness. In practice, this lack of coherence between the aim of the two legislative schemes and the breadth and diversity of content that may fall to be Classified under the OSA, makes classifying online content for the purposes of the OSA consistently across the industry very difficult.

5.5. The lack of clarity around the approach to classification under the OSA also impacts on the development and enforcement of the industry codes and standards that prescribe quite detailed regulatory requirements concerning a range of lawful but harmful Class 1 and Class 2 materials. The development and enforcement of OSA Codes and Standards by reference to the "offensiveness" criteria in the National Classification Scheme is inherently challenging given the vast quantity of content online, much of it being widely dispersed, user-generated and shared by adults in private as well as public online spaces.

5.6. In the case of the Phase 2 Codes, a key objective of the Government is to limit the exposure of children to pornographic materials[26]. DIGI wholeheartedly supports the Government's objective to protect children from exposure to pornography. However, the National Classification Scheme does not define "pornography" as a category of material. Instead, the Scheme sets out a range of contextual criteria which should be considered in determining whether individual instances of material are "unsuitable for children and young people under the age of 18". These criteria are extremely difficult to apply to online materials distributed at scale. In this respect, we note that the research conducted by the eSafety Commissioner on children's access to online pornography, does not use the concepts of the National Classification scheme, but uses an entirely different definition of pornography to evaluate children's experience of pornography online[27]. This underscores the importance of an effective and pragmatic regulatory response to regulating pornographic material that accounts for the risk of harm to children and the realities of the existing online environment.

5.7. DIGI's preference is for the OSA to be decoupled from the National Classification Scheme entirely. We would be very happy to work with policymakers on more effective alternatives. Alternative frameworks might apply a harms-based lens for other categories of materials instead of an approach that centres on "offensiveness"[28]. We also note that the *Roadmap for age*

[25] Ibid.

[26] *Government response to the Roadmap for Age Verification* (Australian Government, August 2023) p.3

[27] Defined as: 'Textual, visual, and audio-visual sexually explicit material that is primarily intended to sexually arouse the audience. This can include representations of images of nudity or semi-nudity, implied sexual activity and actual sexual activity that is uploaded, accessed, and shared via online platforms. This does not include the sending or receiving of nudes or nude selfies, also known as sexting.' See eSafety Research, *Accidental, unsolicited and in your face. Young people's encounters with online pornography: a matter of platform responsibility, education, and choice* and eSafety Research, *Young people, and pornography methodology report* (Office of the eSafety Commissioner, September 2023).

[28] *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography* (Office of eSafety Commissioner, April 2023).

verification, questioned whether the 18 year old age limit for accessing pornography under the National Classification Scheme is appropriate, suggesting that users age 16 and 17 should be able to choose to access online pornography, as they are old enough to consent to sex and are more likely to have received education about sexuality, consent, and respectful relationships by this age. We agree that the issue of the appropriate age limit for pornography should be considered by the reviewer.

5.8. Furthermore, as we noted in section 4 of this submission, the current framework contains a range of confusing and overlapping requirements for the regulation of Class 1 materials in subsidiary regulatory instruments. We suggest that the reviewer consider that rather whether the policy objective of regulating harmful materials could be more effectively achieved by implementing a simplified enforceable standards in the Act that would require services to implement and enforce terms of use that prohibit illegal materials and prohibit/restrict harmful content, proportionate to the service's risk profile. The OSA could also introduce standards that require services to implement and enforce proportionate measures to restrict access of under-age users to certain categories of harmful materials such as pornographic materials.

Question 6 of the Issues Paper asks: To what extent should online safety be managed through a service provider's terms of use?

5.9. DIGI notes that terms of use can be an effective tool that enables services to readily take enforcement action for harmful materials that might otherwise be legally contestable due to differences in the laws of different jurisdictions. For example, services can rely on terms of use, in addition to platform rules or other policies to remove material or suspend/remove accounts of users outside Australia that engage in behaviours that may harm users in Australia. Terms of use can also fill a regulatory gap and can be quickly modified to accommodate types of harmful materials and behaviours that providers identify on their services. For these reasons, the OSA Phase 1 Codes and proposed standards contain mandatory requirements on certain service providers to have and enforce terms of use concerning Class 1 materials.

Question 32 asks: Does Australia have the appropriate governance structures in place to administer Australia's online safety laws?

5.10. In response, DIGI notes that there is a considerable disparity between the administration and governance of offline materials subject to the National Classification Scheme and the administration and governance of online equivalent materials under the OSA. If, the OSA remains coupled to the National Classification scheme then we suggest that the reviewer give consideration to amending the OSA to ensure that regularity parity and consistency between the administration of the scheme offline and online.

5.11. A key finding of the 2020 Steven's *Review of Australian Classification regulation was that:*

> *Classification decisions need to be consistent, accurate, accessible, and easily understood by consumers. The community must have confidence that the right classification outcome is reached, regardless of the process that is used to achieve that classification*[29].

However, the process for classifying materials under the National Classification Scheme and the OSA is very different which is not conducive to promoting consistent outcomes on and offline. The National Classification Scheme provides mechanisms for industry to self-classify content

[29] *Review of Australian Classification regulation Report,* Neville Stevens AO, May 2020 p11.

using classification tools and for content to be submitted to a National Classification Board for review. The Classification Board members are drawn from the wider Australian community and are intended to represent the community and to interpret and apply classification criteria on behalf of the community. In making its decisions, the Board applies benchmarks which provide a level of consistency in classification decisions, while retaining the flexibility to deal with novel content or changes in community standards which may be indicated, for example, through correspondence received, or community or media discussion. Furthermore, under the National Classification Scheme a Classification Review Board has been established to review decisions of the Board. Classification decisions are uploaded and published on the National Classification Database at www.Classification.gov.au with the aim of providing transparent information to Australian consumers and help content providers find content that has previously been classified.

5.12. In contrast, decisions under the OSA are made by a single regulator who can, but is not required to, seek advice from the National Classification Board, without transparency around the decision-making process The OSA approach also effectively requires providers to apply the Classification Scheme at scale, at risk of regulatory action if their approach does not align with that of the eSafety Commissioner This creates a risk of material inconsistencies between the approach to Classification under the two schemes. The Office of the eSafety Commissioner does not provide regulatory guidance or information about the approach adopted by the Office to classifying content online to assist industry with classifying material. Such guidance is crucial given that eSafety and industry participants must take into account the context of material when making classification decisions.

5.13. For example, it is currently unclear how the Office of the eSafety Commissioner approaches the classification of bystander footage of a terrorist act shared by users on a digital service or embedded in a mainstream news story. Under the Classification Act, user generated online video material that depicts a terrorist act is classified pro-terror materials if it advocates for the doing of a terrorist act but is not be classified as pro-terror material if the material can reasonably be considered to depict a terrorist act 'merely as part of public discussion or debate'[30]. This raises the question of whether bystander footage that depicts a terrorist act should be classified as pro-terror material, where a user merely shares the footage online where the material does not contain any element of explicit advocacy, for example.

5.14. Furthermore, the lack of any mechanisms for the recording and publication of classification decisions by the Commissioner make it exceedingly difficult for service providers to determine how they should approach classifying online content under the OSA including under the BOSE, the industry Codes and Standards.

5.15. The lack of any oversight mechanism of classification decisions made by the Commissioner also makes it difficult for the industry participants to challenge decisions by the Commissioner, for example decisions of the Commissioner to issue takedown notices. Indeed, the OSA contains strong incentives on industry participants not to challenge the Commissioner's decisions. For example, an unsuccessful challenge by the provider of a social media service increases the risk of the provider being ordered by the Federal Court to cease providing the service to Australia end-users[31].

[30] Sub-section 9A (3) Classification (publications and Films) Act 1995 (Cth).

[31] See s156 OSA (Federal Court may order a person cease providing a social media service if it contravenes a civil penalty provision in Part 9 twice in 12 months).

Recommendations in Section 5

I. Rather than regulating materials in accordance with the National Classification Scheme, we suggest that the OSA regulate materials that are lawful but harmful in relation to a standard of harmfulness, rather than the current standard of offensiveness.

J. We recommend that the OSA contain clear criteria for categories for harmful materials such as pornographic material that require child specific regulation, such as age restriction.

K. Rather than regulating harmful materials via a range of overlapping and conflicting subsidiary instruments, consideration be given to introducing an enforceable, proportionate standard that services must implement and enforce terms of use that prohibit harmful content, having regard to their risk profile.

L. If the OSA is to continue to rely on the criteria for classifying materials under the National Classification Scheme we recommend that it be amended to:

   a. require the Commissioner to publish clear guidelines for the range of content that falls to be classified under the OSA, with input from the National Classification Board and consultation with academic experts, human rights and civil society representatives and the broader community.

   b. introduce a mechanism to record decisions by the Commissioner to classify material; as Class1 or Class 2, which can be accessed by industry participants and the public.

   c. require independent oversight of the Commissioner's decisions, by an independent board, which represents community interests including establishing an independent mechanism via e persons impacted by decisions of the Commissioner to classify material as Class1 or Class 2 to seek a review of the decisions.

6. The need to strengthen the administration and governance of the BOSE.

6.1. DIGI supports the core principles that underlie the basic online safety expectations. However, **the breadth of the expectations as set out in the current drafting of the BOSE are problematic given the** range of online businesses that are within the scope of the directive.

Question 5: Should the Act have strengthened and enforceable Basic Online Safety Expectations?

6.2. As outlined in section 1, we recommend that the basic regulatory approach be revised by adopting a **proportionate, risk-based based and technologically neutral approach to regulation.** This could include standards that draw upon the common principles that underlie the BOSE, Codes and Standards.

6.3. DIGI has provided detailed input on the scope and requirements of the BOSE in previous submissions[32]. Earlier in this submission we addressed the overlapping regulatory requirements of the OSA, the BOSE, the industry Codes and industry standards. In this section of the submission, we will focus on additional issues related to the administration and governance of the BOSE.

6.4. A core element of the regulatory framework is to 'empower eSafety to seek information from providers on their compliance with the Expectations. This information is sought to improve transparency and accountability, and to assist eSafety determine whether a provider is compliant with the Expectations[33]. Furthermore, this transparency is intended to improve and promote the online safety of Australians by increasing awareness of online safety issues and the way that services respond to online harms. DIGI agrees that these objectives are core to be an effective regulatory framework. With this aim in mind, we recommend the following improvements.

6.4.1. The powers of the Commissioner to request and to publish information under the BOSE notices, extend to commercially sensitive information about providers' safety systems and processes. There is a need for the regulatory framework to provide a mechanism via which providers can seek protection for commercially sensitive information, particularly where providers consider that the publication of sensitive information about their systems and processes may compromise their effectiveness by making them vulnerable to exploitation by bad actors.

6.4.2. The time to respond to BOSE notices is currently 'no shorter than 28 days from the giving of a notice, or from the end of the reporting period specified in the notice". This is too short to enable providers to accurately identify and locate the types of data that are the subject of BOSE requests, particularly where a single BOSE notice covers several different products and services owned by one company. Furthermore, there is no limit on the number of periodic notices with which a provider must comply in any year.

6.4.3. There is a need for additional transparency concerning the Commissioner's regulatory activities under the BOSE in advance of the issue of the annual and periodic notices except in circumstances where there is a need for non-periodic notices to be issued urgently. This would greatly assist providers to respond to the notices in a timely and efficient manner and identify any specific barriers to compliance within the proposed time frame. In particular, the Commissioner should be required to publish an annual plan for the administration of the BOSE that sets out details of:

- the service providers that will be issued with reporting notices (whether periodic or non-periodic) and the criteria for selecting those providers;
- the timetable for the issuing of notices;

[32] DIGI, *Submission on the Online Safety (Basic Online Safety Expectations) Determination 2021* (November 2021) and DIGI, *Submission on revisions to updated Basic Online Safety Expectations 2023* (February 2024)
[33] Office of the eSafety Commissioner, *Basic Online Safety Expectations Regulatory Guidance* (September 2023) p 23.

- the specific criteria that informed the Commissioner's decision that those providers should be issued with notices (e.g. was this based on an assessment of specific risk presented to Australian end-users of those services, or the reach of those services);[34].
- the BOSE expectations that are to be the subject of the notices, and the information that will be requested by the notices; and
- The specific criteria by which the Commissioner will assess compliance.

6.4.4. The BOSE should be amended to include an express right for a provider who has responded to a BOSE notice to be given a draft of eSafety's intended report prior to publication/public comment and an ability to challenge findings before they are made public. The Issues paper emphasises that the BOSE are "not enforceable", yet the BOSE, as noted earlier, is largely treated as quasi-mandatory and providers suffer reputational damage if eSafety forms the view that they have failed to meet expectations with no ability for that to be reversed. The damage is done (and, in effect, the penalty is largely irreversible) once the Commissioner has published/publicly commented on her findings as to whether or not the provider has complied. Given the uncertainty around the BOSE and the way that it has been implemented as a quasi-mandatory regime, this is an extremely difficult position for providers because they need to meet eSafety's specific examples of how they can meet the expectations in the BOSE Guidance, or face reputational damage without effective recourse.

6.4.5. We suggest that the BOSE reporting notices should be focused on requiring the providers to provide an explanation of the steps they are taking to meet relevant measures. Because the BOSE are intended to be flexible, a reporting notice should give a provider a true opportunity to explain the steps they are taking to meet the expectation rather than giving them a limited opportunity to confirm whether or not they are taking a very specific step that is often not expressly mentioned anywhere in the BOSE itself.

6.4.6. Furthermore, providers should be given an opportunity to contextualise the data that they provide to the eSafety Commissioner. As noted above, the specific criteria by which eSafety evaluates compliance should be clear in advance of the issue of notices. However, the types of data that are currently collected do not necessarily produce useful metrics. For example, the number of reports received by a provider in relation to a harm type may increase due to better reporting tools but unless context is given by the provider may be interpreted by the Office as an indicator that the harm is increasing.

[34] We note that to date there are whole sections of the online industry that have not been issued with BOSE notices, including adult websites, dating websites, and chat apps popular amongst culturally and linguistically diverse communities. It is not clear why these types of services have not been subject to scrutiny.

6.4.7. The Office of the eSafety Commissioner has expressed a commitment to taking a consultative approach to the administration of the BOSE. including 'seeking input and feedback from providers as well as from civil society organisations, academics and other experts to ensure implementation meets standards of good regulatory practice'[35]. We recommend that this commitment be formalised, for example by establishing a BOSE advisory Board with which the Commissioner would consult in developing the annual plan recommended in 6.4.3.

Recommendation in Section 6

M. The administration of the BOSE be improved by:

a. Including a mechanism in the OSA via which providers who are the subject of BOSE notices can seek protection for commercial in confidence information.

b. Introducing a requirement in the OSA for the Commissioner to implement and publish an annual plan that sets out its annual program of regulatory activities.

c. Introducing an express right for a provider who has responded to a BOSE notice to be given a draft of the eSafety Commissioner's intended report prior to publication/public comment and an ability to challenge findings before they are made public.

d. Limiting the scope of BOSE notices to requiring providers to provide an explanation of the steps they are taking to meet relevant measures rather than giving them a limited opportunity to confirm whether or not they are taking a specific step that is not expressly mentioned anywhere in the BOSE itself.

e. Establishing a BOSE advisory Board with representation from industry, civil society and experts concerning the annual plan in Recommendation b..

7. The process of drafting industry Codes

Question 5 of The Issues paper asks Should the Act provide greater flexibility around industry codes, including who can draft codes and the harms that can be addressed? How can the code drafting process be improved?

7.1. As outlined in section 1 we consider that the overall flexibility and durability of the OSA could be improved by revising scope and embedding a risk-based proportionate approach to regulation within the Act for groups of services with similar risk profiles/capabilities, rather than the current fragmented approach via subsidiary legislative instruments including industry codes and prescriptive regulatory guidance.

[35] *Basic Online Safety Expectations Regulatory Guidance* p13.

7.2. Setting the recommendation in section 1 aside, DIGI considers that it is important to acknowledge the significant efforts of the Australian online industry in drafting Phase 1 industry Codes, as well as the efforts of the office of the eSafety Commissioner in shepherding the code development process. As a result, codes for six sections of the industry are now in force to address the risks to Australian users posed by Class 1A and Class1 B materials online. Once the Commissioner finalises the Phase 1 standards for relevant electronic services and designated internet services, Australia will have the most comprehensive suite of regulations that tackle seriously harmful materials (including child sexual abuse material and pro-terror material) of any liberal democracy. We consider this a considerable co-regulatory achievement, made possible due to the shared commitment of the office and the many industry participants that contributed to the effort of the code drafting process. The productive working relationship that was established between eSafety and industry in the Class 1 process and the many important learnings we gained should provide a strong foundation for the Class 2 code development process.

7.3. As noted above, we believe that the decision not to register the RES and DIS Codes was a result of flaws in the design of the scheme and could be addressed by reviewing its future scope and providing certainty for low-risk services and those which lack relevant functionality.

7.4. It is also important to acknowledge that drafting co-regulatory Codes was also a novel experience for the Office of the eSafety Commissioner and most industry participants. Consequently, there were significant learnings from the process for both industry and the regulator. As far as industry is concerned, we would note the following challenges, (some of which have been outlined earlier in this submission):

7.4.1. The Commissioner required that Codes be developed for each of the eight sections of the online industry under the OSA. Each code needed to meet a common set of requirements set out in section 141 notices issued by the Commissioner. However, drafting measures for each industry section that could meet these common requirements proved exceedingly complex, due to the diverse functionalities, risk profiles and legal and technical capabilities of services within each industry section. In this respect, the RES and DIS were the broadest and least homogeneous industry sections (see discussion in section 3 above). It was also difficult to identify the appropriate industry representatives who should participate in the drafting process for the catch-all industry section of the DIS. Unsurprisingly, industry was unable to submit Codes that could satisfy the Commissioner's requirements for all the services in scope of the RES and DIS sections.

7.4.2. The OSA Codes require industry participants to determine whether online materials meet the criteria for materials to be refused classification under the National Classification Scheme. This is inherently difficult to do at scale, particularly in the absence of any guidance or database of decisions provided by the Office of the eSafety Commissioner.

7.4.3. The 120 day minimum period allowed for the development of the Codes allowed under the OSA[36] is unrealistic, considering the magnitude of the task and the need to develop 8 codes in parallel which goes significantly beyond what has historically been required under similar co-regulatory schemes operating within

[36] Sub-section 141 (2) OSA.

the broadcasting and telecommunications industries.. The section 141 notices issued by the Office of the eSafety Commissioner allowed industry 6 months to complete the development of the Codes, but this was also insufficient: the Code development process for the Phase 1 Codes in fact took 20 months. The short time allowed for code development under the OSA, meant that for much of the code development process the industry was operating under a high degree of uncertainty as to whether it would be able to deliver the Codes 'on time'. In DIGI's view, a period of 12 to 18 months would be a more reasonable timeframe for the development of future industry codes.

7.4.4. Industry relied on the Position paper for Code development published by the Commissioner to guide the development of measures which contained examples of how industry could meet the Commissioner's requirements for registration (noting the very broad discretion the OSA affords the Commissioner to decide whether or not to register a Code). As the drafting process unfolded, a considerable gap emerged between the expectations of the Commissioner and industry on the specific measures that should be implemented for different industry sections to proactively detect Class 1 Materials. This led to prolonged debates between industry and the Office, concerning the extent these expectations were realistic and achievable. In retrospect, this could have been addressed to some extent for more extensive consultation with industry and other key stakeholders so that both sides had a common understanding of the Commissioner's expectations about the 'non-negotiable' key measures that should be included in each Codes.

7.4.5. While the knowledge gained by the Office and industry participants from the Phase 1 Code drafting process will no doubt benefit the drafting of Class 2 Codes, the fundamental issues that we have identified with the approach of the OSA to categorising services and regulating harmful but lawful materials and the limited time allowed by the OSA for codes development remain. DIGI therefore recommends that in order to produce optimal regulatory outcomes, the Class 2 process be deferred until these fundamental issues have been considered by the reviewer and any relevant recommendations considered by the Government. Alternatively, we ask the reviewer to recommend that at least 12 months is allowed for these Class2 Codes to be developed.

Recommendation in Section 7

N. We recommend that the Class 2 OSA Codes development process be deferred until the OSA review concludes and any relevant recommendations have been considered by the Government. Alternatively, we ask the reviewer to consider recommending that at least 12 months is allowed for these Class 2 Codes to be developed.

8. The eSafety Commissioner's enforcement and Investigative powers

8.1. Part 4 of the Issues Paper raises questions about whether the penalties for breaches of the OSA are adequate and whether the eSafety Commissioner enforcement and investigative powers should be expanded. On the basis of comparisons with international online safety legislative regimes, the Issues Paper seems to assume that the answer to these questions must be in the affirmative: "Broadly speaking however, Australia's penalties regime has not kept pace with newer regulatory regimes, such as in Ireland, the EU, and the UK, which apply significantly higher penalties, including penalties based on a percentage of a platform's global revenue".[37] We do not think these comparisons are necessarily useful for assessing the effectiveness of the penalties and investigative powers of the Commissioner under the OSA . Many of the international examples referenced in the *Issues Paper* are relatively new and are based on very different regulatory approaches. Further, as DIGI has identified throughout this submission, the OSA has a much broader scope than the comparative overseas regimes (both in terms of the industry participants and the types of content which is regulated). The Commissioner also has range of powers under the OSA that are not available to overseas regulators e.g. the power to delist an app from an app marketplace, powers to engage in adverse publicity under the BOSE, and the power to seek an order in the federal court that a person cease providing a social media service if it contravenes a civil penalty provision in Part 9 of the OSA twice in 12 months).

Recommendations in Section 8

O. Any proposed changes to the Commissioner's regulatory and investigatory powers should be based on a thorough evaluation of effectiveness of the current enforcement and penalty regime, and should be primarily directed at addressing identified deficiencies with the effectiveness of the current regime

[37] *Issues Paper* p.33.

3. Concluding Remarks

In conclusion DIGI would like to reinforce our commitment to online safety regulation that foster equitable safe digital spaces for the Australian community. DIGI sees itself as a key Government partner in these endeavours. We hold a long track record of directly engaging in the development of regulation that is effective in its goals and can practically be implemented by industry. We appreciate the opportunity to contribute to this process and look forward to the outcome of this review.

APPENDIX

The statutory review (review) of the Online Safety Act 2021 (OSA) occurs against the background of no fewer than fifteen other related reform or consultation processes. Those are:

1. the recent Online Safety (Basic Online Safety Expectations) Amendment Determination.
2. 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).
3. the recent Government announcement of a pilot of age assurance technology.
4. the second stage of the modernisation of Australia's National Classification Scheme.
5. the development of industry codes for Class 2 material.
6. processes for the making of necessary subordinate regulation (rules/standards) as a consequence of the recently passed Digital ID Bill 2024, together with the Digital ID (Transitional and Consequential Provisions) Bill 2024.
7. the review of the Privacy Act 1988 (Privacy Act) (with a foreshadowed introduction into Parliament in August 2024), including Government's agreement to implement a Children'sOnline Privacy Code to promote the design of certain services in the 'best interests of the child'.
8. foreshadowed legislation addressing hate speech and religious discrimination.
9. the Misinformation and Disinformation Bill 2023 and associated processes.
10. the voluntary code for online dating services.
11. the recently established Joint Select Committee on social media and Australian Society.
12. the establishment of the Select Committee on Adopting Artificial Intelligence, which will inquire into, among other things, the risks and harms arising from the adoption of AI technologies, and emerging international approaches to mitigate AI risks.
13. 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 and potential mandatory requirements for high-risk use cases.
14. the anticipated Government response in relation to dispute and complaints resolution processes of digital platforms, flowing from the ACCC's Digital Platform Inquiry; and

15. the Department of Home Affairs report in relation to understanding algorithms on digital platforms.