Microsoft submission to the **Statutory Review of the Online Safety Act 2021**

Submission to the consultation process run by the **Department of Infrastructure, Transport, Regional Development, Communications and the Arts**

21 June 2024

1 Introduction

Microsoft welcomes the opportunity to respond to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (the **Department**) consultation in relation to the statutory review (the **Review**) of the *Online Safety Act 2021* (Cth) (**OSA**). We have prepared this submission based on the *Statutory Review of the Online Safety Act 2021 – Issues paper* (the **Issues Paper**) as well our experience working closely with eSafety and other local stakeholders in relation to the OSA. Our submission includes the following observations and recommendations:

1. We recommend taking advantage of the opportunity afforded by the Review to:
    a. simplify and streamline the OSA;
    b. align the OSA with emerging international best-practice; and
    c. build community trust by increasing transparency and safeguarding human rights across the online and regulatory ecosystem.
2. We support an enhanced regime centred on either a statutory duty of care and an improved focus on systems-based regulation.
3. We offer targeted, pragmatic feedback on existing aspects of the OSA, including:
    a. the Basic Online Safety Expectations;
    b. the Online Content Scheme, including Industry Codes and Industry Standards; and
    c. the clarity of service categorisations and definitions.

2 Microsoft's approach to digital safety

Microsoft recognises the unique role that technology companies play in helping make the internet safer, as well as our responsibility to reduce the risk of harm from content or conduct on our services, especially for children. We take steps to address illegal and harmful online content across our services,

while balancing our commitment to respect fundamental rights such as privacy, freedom of expression, and access to information. We do this through a risk proportionate approach: tailoring our safety interventions to the nature of the risk and the unique characteristics of the online service. We take steps to advance safety across four interconnected pillars:

- Platform architecture: Advancing safety be design in the development and operation of our services, including through advancing internal standards, considering risks, and building in safeguards such as family safety tools.

- Content moderation: Ensuring we have clear, consistent and transparent processes to address content and conduct that violates our policies.

- Culture: Empowering users to foster safe online spaces and building an understanding of the risks and opportunities of life online.

- Collaboration: Working closely with a wide range of stakeholders to address complex, whole-of-society digital safety challenges and hear insights from diverse perspectives.

Microsoft has supported the development of legislative measures as a part of an effective whole-of-society approach to digital safety, where such measures are clear, practical, proportionate, and enable an effective approach to online harms. Based on our experiences since the Online Safety Act was enacted in 2021, we provide suggestions to refine this regime to provide additional clarity and proportionality.

3 Opportunities for an improved OSA

The OSA's comprehensive regulatory framework speaks to the strength of the Australian Government's commitment to digital safety. While the OSA is pioneering in many respects and has had considerable global influence, the Review presents the Government with an opportunity to improve its current operation and effectiveness, as well as its sustainability over the long-term.

3.1 Simplified and streamlined regulation

The online harms impacting users around the world are complex and constantly evolving. Responding to these harms effectively and proportionally is equally complex and requires carefully calibrated regulation. The scope of harmful content and conduct online (ranging from offensive exchanges to the most serious of criminal activity) means that regulation must balance broad, adaptable principles with requirements that are targeted and clear. Overly complex regulation not only frustrates compliance by participants in the large and diverse 'online ecosystem' but also undermines the enforceability of protections, public awareness and the digital safety of individuals.

While its wide scope is part of its strength, the OSA can be a complex and convoluted regulatory framework to navigate. This is contributed to by the following factors:

- **Spread of obligations**: The OSA includes detailed frameworks governing each of the image-based abuse, cyberbullying and cyber-abuse schemes, and the Online Content Scheme. In addition, the OSA provides the foundation for additional regimes via both delegated regulation, such as in the case of the Basic Online Safety Expectations (**BOSE**), and co-regulation, in the case of industry codes under the Online Content Scheme (**Industry Codes**), as well as the industry standards under the Online Content Scheme (**Industry Standards**).

These requirements range from conventional notice-and-takedown requirements to service-specific, risk-based obligations under the Industry Codes, to principles-based expectations under the BOSE. Further, the OSA is not the only relevant Australian law when it comes to digital safety. Local privacy, telecommunications, defamation, electoral, consumer, discrimination and criminal laws, among others, have significant digital safety implications on service providers, as will forthcoming legislation in relation to misinformation and disinformation.

- While intended to provide an end-to-end approach to advancing safety across the online ecosystem, the multiple, overlapping regulatory regimes (both within the OSA framework and beyond) has created significant challenges for in-scope services to map their obligations in Australia and determine how to prioritise action and understand what compliance requires. For instance, as outlined in our submission to eSafety on the draft Industry Standards, the Head Terms for the Industry Codes was not adopted in its entirety for the draft Standards, creating different frameworks even within the Online Content Scheme. While legislative compliance is of course a cost of doing business, the current complexity increases this significantly, especially for companies looking to diversify their service offerings. The current regime also creates a risk of duplicative obligations (e.g., for transparency reporting) and so could benefit from streamlining, consolidation, and simplification.

- **Legislative foundations**: The OSA reflects a significant uplift and amalgamation of several directly preceding statutory schemes, namely the *Enhancing Online Safety Act 2015* (Cth) and schedules 5 and 7 of the *Broadcasting Services Act 1992* (Cth). By borrowing and modernising concepts from previous statutes, parts of the OSA fail to properly capture the nature of the contemporary internet or do so in a convoluted way.

  In addition, the OSA derives its conception of offensiveness, and of class 1 and class 2 material, from the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) and subordinate instruments (together, the **National Classification Scheme**). In addition to being outdated and undergoing its own process of review, as a regulatory framework primarily designed for the classification of professionally produced, legacy media formats, the National Classification Scheme is ill-equipped to respond to the nature and scale of online content, particularly user-generated content, which is both high volume and often highly context-specific. The offense-based approach of the National Classification Scheme is also difficult to reconcile with the risk-based and harm-based approaches most suitable for effective digital safety regulation, especially when reconciling user protection with human rights at scale.

  The misalignment between the National Classification Scheme and the OSA has been widely observed, including through the 2020 Review of Australian Classification Regulation undertaken by Mr Neville Stevens AO (the **Stevens Review**) and subsequent policy reform fora. As we explore further in this submission, a continuing link between the National Classification Scheme and the OSA may be tenable, however, the two should be made more interoperable than they are currently. This concurrent review of both schemes presents a unique opportunity for this harmonisation.

- **Lack of clarity on scope:** Leveraging the National Classification regime for the Online Content Scheme has also contributed to a significant lack of clarity on the kinds of illegal and harmful online content to be addressed under the OSA. As a matter of principle, Microsoft has consistently advocated for legislation that clearly defines any in-scope content, and that draws bright lines between illegal content and content that is lawful but harmful. The latter should be left to companies to address through their terms of service, enabling an effective

approach to address harm and avoiding legal requirements to remove content that has not been outlawed through a democratic legislative process.

- The principal OSA does this effectively, through the removal orders scheme for certain, specified categories of content, but this clarity breaks down across the totality of the regime, which risks undermining the speech rights of Australians, as well as the effectiveness of the overall regime. In addition to the challenges noted in our prior point, this is exacerbated by the breadth of content in scope for the BOSE. Regulated services in Australia therefore risk being unable to effectively understand their compliance obligations, to prioritize compliance measures, and over-moderating content from Australian end-users. It also makes it challenging for the Australian public to understand how they may be affected by these measures.

- **Constant development and reform**: As digital safety concerns evolve, the capacity for existing regulations to respond is often called into question. Indeed, it can be said that digital safety laws in Australia have been in a constant state of flux since 2018. Such activity has taken various shapes during this period:

    - formal legislative reform processes (i.e. the transition from the EOSA to the OSA);

    - subsequent development processes of regulation subordinate to the OSA (i.e. the development of Industry Codes and recent reforms to the BOSE);

    - proposed additional reforms (i.e. the *Combatting Misinformation and Disinformation Bill 2023* and recent anti-doxing consultations), including those that were ultimately abandoned (i.e. the *Social Media (Anti-Trolling) Bill 2022*); and

    - recurrent dialogue in the political sphere and developments through technology pilots, parliamentary select committees and other investigation and recommendations processes, and broader community for new, tougher laws.

  Regular review of regulation to ensure its fitness for purpose and overall activity in this space underscores a healthy digital safety conversation in Australia. However, there is an extent to which the continuation and pace of reform can undermine the effectiveness of compliance efforts and positive outcomes for Australians. Sometimes, the pace of reform has meant that potentially fruitful regulatory schemes are not afforded a sufficient 'runway' for compliance to establish and mature organically.

  Microsoft believes that the effective operation and longevity of the OSA could be bolstered through changes to the structure and design of regulation, focusing on clear, enforceable and tech-agnostic principles that can adapt to changing settings. Designing regulation around the systems and processes implemented by service providers can provide clarity for implementation and enable flexibility to address the evolution of technology and the online harm environment.

3.2 Future-focused, globally coherent

Since the OSA was developed and introduced, there has been significant global progress in digital safety regulation, including the UK's Online Safety Act 2023 (**UKOSA**) and the EU's Digital Services Act (**DSA**). The Review presents an opportunity to consider evolving global best-practice and increase the alignment of the OSA with international digital safety regulation.

As recently noted by the Global Online Safety Regulators Network (**GOSRN**), (of which eSafety is a founding member) "…neither the risks people face online, nor the online services they use, are confined to national or continental borders". While this does not diminish the role of domestic regulation, it does create an incentive to adopt coherent regulatory frameworks. Coherence can better enable service providers to respond to online harms more effectively from an outcomes perspective, and more efficiently from a resource perspective. Importantly, it also fosters a safer digital environment for all users (regardless of their location), promotes legal certainty, helps uphold fundamental human rights and improves community understanding.

From an enforcement perspective, reforming the OSA in line with international regulation better allows Australia to leverage collective influence and shared expertise to tackle common online harms, and provides an opportunity to develop unified regulatory tools and practices. For example, coordinated information requests have the potential to create a more predictable regulatory environment, benefiting both regulators and the industry.

As we touch on in this submission, the UKOSA and DSA each approach digital safety at a more systemic level than the OSA presently does. We argue that such approaches would make Australian digital safety regulation better placed to adapt to the rapid pace of technological change and patterns of harm.

3.3 Transparency, rights and trust

Microsoft acknowledges the important role that transparency has in boosting accountability and compliance with digital safety regulations. We are also fierce advocates for regulation that safeguards fundamental human rights and recognise the need to balance digital safety with privacy, freedom of expression or equal access to online services.

In order to support a relationship of trust with our users, we provide public-facing information on our digital safety policies and practices. We also publish a range of transparency reports (available on Microsoft's Reports Hub) across our business in respect of a variety of issues including content removal requests, law enforcement requests and our bi-annual **Digital Safety Content Report**. In addition, we acknowledge the unique issues, risks and harms that arise from different products across Microsoft's offering and release dedicated transparency reporting for Xbox, LinkedIn and GitHub. In recent years, we have also published an Australia-specific report on our efforts to combat mis- and dis-information on relevant services under the voluntary code of practice. Microsoft has made human rights commitments, outlined in our Human Rights Statement, and provides human rights reporting.

Both transparency and rights-based safeguards are key to engendering community trust in this context. While the providers of online services are the primary subject of digital safety regulation, it is individuals and online communities whose activities are ultimately regulated in the effort to strengthen online safety. Around the world, including in Australia, there are varying societal attitudes toward regulation that intervenes in the online world. This directly impacts not only the trust individual users have in their online service providers, but also trust in government and other official institutions.

Above all, carefully navigating this community trust is critical to the effectiveness of digital safety regulation. Microsoft believe that there is an opportunity to both increase transparency and bolster human rights under the OSA and in its enforcement processes.

- **Meaningful transparency**: Transparency acts as cornerstone of several digital safety regulations around the world, and Microsoft support its continued role in the OSA.

  In this submission we provide targeted feedback in relation to how transparency reporting currently operates via the BOSE, but also include the following views on best-practice transparency:

  - To the extent possible, transparency reporting requirements within the same regulatory area should avoid duplication and aim for coordination with reporting under substantially aligned regimes. Accordingly, there is an opportunity in revising the Act to support mutual recognition across international regulatory regimes, whereby the eSafety Commissioner may be empowered to accept transparency reporting under a comparable safety regime or formal requests for information from another regulator as adequate for the purposes of Australian regulatory compliance. This would support global efforts to drive regulatory coherence and uplift safety practices. For example, Microsoft understand that it is a goal of the GOSRN to explore better coordination on information requests to support more globally comparable data and reduce compliance burden for service providers.

  - Transparency mechanisms work most effectively when they are used (at least at first instance) as a cooperative tool to increase a regulator's understanding of industry activities. While service providers who are uncooperative or obstructive should be held accountable, there are opportunities to enhance safety through a collaborative approach.

- **Regulatory transparency**: Just as transparency from service providers helps foster trust and accountability, greater transparency of regulator enforcement activity of the regulator is critical to building wider community trust in digital safety regulation.

  As is the case with certain information shared with the eSafety Commissioner by service providers, we recognise that not all enforcement-related information is suitable for public release. We also note the application of federal freedom of information requests to the eSafety Commissioner. The OSA should consider clear processes for the regular and proactive disclosure by the eSafety Commissioner of the exercise of its powers, both formal and informal.

  Clearer reporting around the exercise of both informal and formal regulatory powers can help codify a record of relevant decision-making in Australia, benefitting industry compliance and community awareness alike. Improved clarity in this regard could also be beneficial for the protection of human rights, as service providers that may otherwise adopt an overly censorial approach due to ambiguity will have clearer precedent for decision making. We also foresee greater transparency having positive outcomes for local technology innovation and investment by reducing regulatory unpredictability.

  Rights-based protections: Both the UKOSA and the DSA expressly contemplate service providers appropriately and proportionality balancing digital safety interventions with user privacy, security and freedom of expression. The OSA has no such protections, and has attracted widespread commentary from civil society, academia and industry on its active potential to undermine the fundamental human rights of Australians.

  Human rights considerations are particularly exposed in the OSA's subordinate instruments, such as the Industry Codes or the BOSE. It is not enough, for example, for human rights to

merely be considered as part of co-development processes or the development of instruments like the BOSE. Without requirements hardcoded into the body of the OSA, service providers (through their compliance efforts) and the eSafety Commissioner (through its enforcement) may interact with broadly drafted requirements in a manner that impacts fundamental human rights. We therefore recommend the review takes the opportunity to consider how human rights may be better coded into the OSA as primary regulation that will then flow down into the secondary mechanisms.

4 Commentary on best-practice regulation

4.1 Statutory duty of care

As discussed above, Microsoft views the complexity of the OSA, and its ongoing supplementation through delegated legislation and reform processes, as having the potential to result in sub-optimal digital safety outcomes and progress for Australians. This, and a desire to see greater international harmonisation, is why Microsoft supports the Department exploring how a carefully framed statutory duty of care for online service providers could be incorporated into the OSA.

A statutory duty of care could clearly demonstrate the Australian Government's expectations for providers of the online services that are accessed by Australians, in a manner that can flex to technology innovation and an evolving online harm landscape. The introduction of a duty of care, together with clearly defined harms and a systems and processes-based approach to regulation, carries the ability to adapt as necessary as technology and community behaviours change. Specificity on how to discharge that duty of care could then be highlighted through guidance and related materials produced by eSafety (e.g., by adapting the existing codes regime). Taking this approach would also be an opportunity to ensure the OSA is advancing a risk-based and proportionate approach, tailored to the unique nature of the services in scope.

Incorporation into the OSA: Care should be taken to ensure that any statutory duty of care incorporated into the OSA is clear and principles-based, enabling service providers to apply the principles as appropriate and proportionate to their specific platform and services. For example, in the UK context, while codes and guidance from the UK regulator Ofcom offer evidence-based examples of expected compliance measures, in-scope services may nonetheless take different approaches.

As with other duties of care, effectiveness is linked to the ability for those who owe the duty to discharge it in in a flexible and context-appropriate manner. In exploring any incorporation of a duty of care into the OSA, the Department should carefully consider how such a duty interacts with other existing or additional legislative requirements, being careful not to unnecessarily blunt the effectiveness of the duty with overlapping, conflicting or duplicative obligations.

The Issues Paper noted that the *2022 House of Representatives Select Committee on Social Media and Online Safety* had supported introducing a formal statutory duty of care framework as an enhancement to the BOSE. While we acknowledge the relationship between a duty of care approach and the BOSE, we caution the Department against viewing the BOSE as a suitable vehicle for introducing a duty of care into the OSA. Aspects of the BOSE, particularly as recently amended, are simultaneously too broad and too prescriptive to form the basis of a duty of care, the benefit of which lies in its principles-based and adaptable approach. Additionally, a statutory duty of care should be drafted into the text of the OSA, so that amendments go through the legislative process rather than ministerial instrument. Introducing a duty of care approach would also require re-evaluation of the role of the BOSE and Online Content Scheme, presenting an opportunity to streamline and simplify the overall regime.

4.2 Greater focus on principles, and systems and processes

Even where a statutory duty of care is not incorporated into the OSA, existing schemes, including the BOSE and active Industry Codes, could nonetheless benefit from a greater focus on adaptable principles, and a systems and processes approach to harm.

Due to the great diversity of online services and their respective user communities, service providers need flexibility in how they comply with digital safety regulation to adequately safeguard fundamental human rights. Drawing on many of the same observations as noted above in relation to a duty of care, the OSA's regulatory schemes function at their best when they focus on systems and processes, and adaptable principles rather than prescriptive requirements. The current Industry Code and Industry Standard development processes carry the potential to produce overly prescriptive obligations with a shorter shelf-life than desirable. Any changes to the OSA regime should focus on being technology neutral and outcomes-based.

5 Specific commentary on existing regime

5.1 Basic Online Safety Expectations (BOSE)

General: The BOSE are described as setting out the Australian Government's minimum safety expectations of online service providers. The BOSE apply to all services falling within the broad categories of social media service, relevant electronic service, or designated internet service. Due to not only the different types of in-scope services covered by the BOSE, but also the different harms that arise in respect of each type of service, it is important that the BOSE can accommodate this diversity and recognise that expectations will have varying levels of relevance for different services. Alignment with the BOSE may also look significantly different depending on the service and its risk profile, even within the three broad categories of in-scope services. If retained under the OSA regime, Microsoft believes that the BOSE would operate most effectively by prioritising a principles-based approach that offers service provides the flexibility and self-determination to choose how they foster safety on their services.

It is also important that the flexible application of the BOSE be kept in mind in its enforcement by the eSafety Commissioner via reporting and transparency notices. By design, the BOSE proposes benchmarks that do not account for differences in size, risk profiles and user thresholds across service providers. To date, the eSafety Commissioner has issued BOSE transparency notices primarily in multi-provider tranches, with associated public reporting on each provider's response being made together with each other provider in that tranche. The comparative nature of this approach may align with aims to collectively lift industry standards, but can often ignore the considerable differences between providers. As it stands, the broad scope and prescriptive qualities of the BOSE mean it sits uneasily with the harms expressly addressed in the OSA nor with the intention for a risk-based framework through the Online Content Regime.

Transparency, confidentiality and safety: Transparency for both online service safety interventions and regulatory enforcement can provide important information to governments, civil society, and end-users to contextualise, understand, and advance online safety outcomes. Microsoft recognizes that the imperative for transparency must be balanced with competing digital safety, cybersecurity, privacy and confidentiality interests. For example, certain disclosures can provide bad actors with knowledge about a provider's efforts to combat malicious activities and allow the exploitation of safety controls and processes, subsequently increasing the risks of online harms rather than mitigating them.

Currently the OSA does not expressly contemplate the reporting under the BOSE being set for public release, but there is no statutory requirement for the eSafety Commissioner to (a) consider applications from service providers to keep information confidential, or (b) balance any competing considerations when disclosing information to the public. We recommend the Department consider the need for conditionality to be introduced to the OSA in order to balance public disclosure with competing interests.

5.2 Online Content Scheme

Interaction between OSA and the National Classification Scheme: The Online Content Scheme contained in the OSA seeks to limit harms associated with specific forms of online content known as "class 1 material" and "class 2 material". These two classes encompass a broad range of illegal material (such as child sexual exploitation material), as well as restricted material (such as online pornography). As outline above, the OSA derives its conception of class 1 and class 2 material from the National Classification Scheme. Even once the National Classification Scheme is amended there is still a legitimate question as to whether classification laws are appropriate for the regulation of online material. While some degree of linkage between the National Classification Scheme and the OSA may be desirable, amendments could be made to each regime so that relevant concepts are operating in a more intuitive and appropriate way.

Industry Codes and Industry Standards: In line with comments made in relation to improved transparency and human rights protections within the OSA, Microsoft considers there is an opportunity to improve the existing Industry Codes and Industry Standards mechanisms. As outlined above, one approach may be to revisit the function of the Industry Codes and Industry Standards to become guidance in steps that could be taken to discharge a duty of care, rather than as imposing additive mandatory obligations.

With respect to transparency, we note that despite Industry Codes being led by service providers and industry associations, in practice the process has been driven by detailed expectations from the eSafety Commissioner, who also has the final say over whether Industry Codes will be registered. As such, much of the development process involves iterative and private negotiation between industry and the eSafety Commissioner. Given the potential scope and impact of such regulation, and notwithstanding mandated public consultation windows, it is worth considering whether any increased transparency is required over such deliberations. In addition, the OSA could benefit from better establishing the respective roles of industry, industry associations, and the eSafety Commissioner when it comes to Industry Codes, as well as whether such potentially consequential regulation ought to go through either independent review or some form of parliamentary scrutiny.

There are presently few statutory guardrails placed on either industry or the eSafety Commissioner, in terms of what these forms of subordinate regulation can regulate. Not only does this add to protracted development processes for such regulation, but also risks the formalisation of regulation that goes beyond the scope of the OSA as agreed by Parliament. And, as outlined earlier, protection for human rights is not currently encoded in that regime.

Finally, given the current role that the concept of "community safeguards" plays in determinations by eSafety of whether to register Industry Codes, consideration should be given to defining that term in the OSA itself, along with further transparent and objective criteria to guide the Commissioner's decisions

5.3 Clarity of service categorisations and definitions

Service categorisation: The way in which the OSA categorises online services, and their associated definitions, may risk needlessly complicating the regime. This complexity has the potential to undermine compliance by service providers and, by extension, the general effectiveness of the OSA.

While the definition of 'social media service' provides a reasonably clear path to determining where a service is a social media service, the vagueness and overlapping nature of the 'relevant electronic service' and 'designated internet service' categories means that there is significant confusion around how to categorise services that do not fit cleanly into any one category. These category names are not intuitive and do not provide service providers or consumers with any insight into what types of services may be covered by these categories. Further, the vagueness also impacts the public perception and understanding of the OSA regime and what rights users have in respect of their online safety. We therefore recommend giving consideration to better tailoring the definitions in the Act to enable the more appropriate application of targeted safety measures, coupled with a risk-based approach.

Designated internet service: The definition of designated internet service results in essentially all websites and other online services accessible in Australia being captured by this definition, and therefore subject to the OSA. While there was a degree of intentionality behind this, the breadth of the service category leads to numerous compliance challenges in practice.

For example, as the BOSE apply to all designated internet services in the same way and without any materiality or metric-based thresholds, personal blogs run by individuals face the same obligations as large corporate websites and online services. In an Industry Codes context, such a broad service category also frustrates the efficient development of coregulation with significant sub-categories of the designated internet services section being required and making the codes unworkably complex.

Microsoft recommends that the definition of designated internet service is clarified to focus on services that represent real or probable harm and excluding general purpose websites. Excluded websites could include websites operated for lawful trade or commerce by small businesses, websites used for personal or domestic purposes, or services without a user-to-user interactivity.

6 Conclusion

Microsoft appreciates the opportunity to provide input to the Australian Government's review of the OSA. As outlined in our submission, we feel this is an opportune time to simplify, streamline and harmonise the OSA regime to help preserve its flexibility to tackle emerging online harms. We offer some suggestions and observations to help the government achieve its vision of effectively confronting online harms while respecting fundamental rights and freedoms, and applying a proportionate approach to regulation.

We would very much appreciate the opportunity to discuss our submission directly with the independent reviewer and provide clarity on our recommendations.