To: Joint Select Committee on Social Media and Australian Society,
PO Box 6100, Parliament House, Canberra ACT 2600

By email: socialmedia.joint@aph.gov.au

Wednesday 3 July 2024,

Dear Ms Thwaites MP,

The Digital Industry Group Inc. (DIGI) thanks the Government for the opportunity to provide our views to the Joint Select Committee on Social Media and Australian Society.

DIGI shares the Government's commitment to improving online safety for Australians, which is a central pillar of DIGI's organisational mission. In driving forward our vision of a thriving Australian digitally-enabled economy that fosters innovation where online safety, privacy and consumers are protected, DIGI collaborates with the digital industry, Government and community stakeholders. Our work spans expert advocacy for effective and implementable approaches to technology policy, code development, and partnerships. Our work involves working closely and collaboratively with the world's leading technology companies who are also committed to this vision, and our founding members are Apple, eBay, Discord, Google, Linktree, Meta, Microsoft, Snap, Spotify, TikTok, Twitch, X, and Yahoo.

DIGI believes that all digital technology – including social media services – should be safe, equitable and hold a net positive influence on society, and that any negative impacts must be actively mitigated. DIGI, and its members, unequivocally believe that the digital industry has a responsibility to address online harm. From our unique vantage point, DIGI is in a position to see the extensive, multi-faceted work our members undertake in addressing online harms, data protection and consumer protection. The challenges are not simple, so industry investment in continued improvement to effectively address existing and emerging threats online is critically important. In addition, the Australian Government has an important role to play in standardising protections, encouraging corporate accountability and providing safety nets for consumers across the industry, as well as cross-sectorally in a digital economy.

For our part, with the full support of our members, DIGI plays an important role in relation to improving and standardising protections for Australians online. For example**:**

- DIGI co-led the development of mandatory codes required under the Online Safety Act, that provide Australians with new enforceable protections against child sexual exploitation material, pro-terror content and other extremely harmful materials (Class 1 codes). This involved extensive and iterative engagement with the Office of the eSafety Commissioner and a wide range of industry participants to develop binding and implementable systems and processes to address seriously harmful 'Class1' materials online (being categories of materials that would be refused classification under the National Classification Scheme). Six of these codes are now in force, creating new online harms legal obligations for social media services, app distribution services, search engines, equipment providers and related services, hosting services and internet service providers. DIGI has also engaged extensively with the Office of eSafety Commissioner on the development of their draft standards for Designated Internet Services and Relevant Electronic Services, which were registered on 21 June 2024, and will come into effect on 21 December 2024.
- At time of writing, DIGI is currently consulting with the Office of the eSafety Commissioner concerning the development of Phase 2 industry codes to regulate materials that are unsuitable for children and young people under 18 years old under the National Classification Scheme.

- DIGI also works on the prevention of online harms, and improving young people's digital literacy and capacity online, through our DIGI Engage youth summits. Over multiple years, DIGI partnered with the

Australian Government – through the Department of Home Affairs, Multicultural NSW and the Attorney General's Department – around our shared goals in relation to countering violent extremism. DIGI Engage youth summits have upskilled hundreds of young people about the root causes of societal polarisation, hate speech and extremism, and has built their capability online and offline to counter them. In 2024, DIGI has a current partnership with the NSW Government's COMPACT Program, which will bring together practitioners and young changemakers working to promote social cohesion in their communities and counter hate online.

- DIGI developed and oversees *The Australian Code of Practice on Disinformation and Misinformation* (ACPDM). Since its launch in February 2021, DIGI has made continued improvements to the code, including the introduction of a public complaints portal for breaches of the code, the appointment of independent complaints committee, independent attestation and assessment of signatories' annual transparency reports, and a public review of the code which included public consultation. DIGI supports the creation of a regulatory backstop to the code through increased powers for the Australian Communications & Media Authority (ACMA), and we will continue to partner with the Government, the ACMA, and signatories to seek improvements for the code.

- DIGI is engaged in multiple working groups in the National Anti-Scam Centre (NASC), including its advisory board, recognising the responsibility of the digital industry – including social media services – as part of a cross-sectoral ecosystem approach to protect Australians from scams. DIGI has been working with an industry working group to consider what implementable industry level obligations might look like in the diverse digital industry, as a complement to, and in the context of the Government's legislative agenda in relation to scams.

DIGI welcomes this inquiry as a way to have a deeper conversation about online safety. The online environment is always evolving, as are platforms' online safety measures, and there is value in considering and assessing these conditions to aid continual improvement. Continued dialogue is extremely important, as online safety can never be a set and forget exercise.

DIGI's work in the above areas and others is a reflection of our members' deep commitment, longstanding and continued investment in online safety. In this submission we draw on our relevant experience to support the Government's assessment of the matters outlined in the Inquiry's terms of reference. We hope that this submission assists in advancing a shared understanding of the current landscape, themes and potential gaps. Online harms are multi-faceted social problems that cannot be fixed with technical and legal safeguards alone; this is why we are a proponent of multi-stakeholder approaches in relation to online harms that continue to ensure strong accountability and responsibility on the part of online platforms, while also situating platform-level responses in a wider context.

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

Best regards,

Sunita Bose
Managing Director
Digital Industry Group Inc. (DIGI)

Table of contents

1. Social media in Australian society

1.1. Social media has diverse audiences and uses

1.2. The Government has noted that social media is how millions of Australians connect, access information and run small businesses.[1] DIGI agrees that social media services[2], and how audiences use them, are multifaceted. For example, the types of activities and experiences Australians might be engaging in when using a social media service include:

1.2.1. Communication and community building (e.g. staying in touch with friends and family around the world, reading updates in local neighbourhood groups or connecting with groups around common interests)

1.2.2. Accessing information and educational resources (e.g. digital resources of a Museum collection, a professional network, real time health and safety updates from emergency services or health authorities, microcredential learning)

1.2.3. Running a business (e.g marketing and advertising services, customer service and engagement, running a digital shopfront)

1.2.4. Civic engagement and social activism (e.g charity fundraising, social awareness campaigns, digital petitions, engaging in public debate)

1.2.5. Entertainment (creating and sharing content/user generated content for entertainment purposes in the realms of fashion, games, sport, arts and culture)

1.3. We outline these categories to demonstrate that, while we might use social media as a generalised term through this submission, there are a number of activities and behaviours that might be undertaken when an Australian uses a social media service. The multitude of uses for social media is important to acknowledge when assessing its potential influence on consumers and society as a whole, and when considering the implications of policy proposals.

2. Industry and Government approaches to reduce risk of online harms

2.1. DIGI and its members recognise that the digital industry has a crucially important responsibility to address online harm. Working from a shared understanding of the extensive work undertaken by industry, and underpinned by regulation, is foundational to the committee identifying where there may be gaps that it wishes to explore.

2.2. DIGI has detailed in (Appendix 1, p.17) generalised trends in the digital industry in relation to online harms relevant to the terms of reference of this inquiry. At a high level, there is a long-standing and continued investment in improving online safety outcomes through policies that define what content and behaviour is permitted on their service and what is

[1] May 2024, Albanese Government's Joint Select Committee into social media, https://ministers.treasury.gov.au/ministers/stephen-jones-2022/media-releases/albanese-governments-joint-select-committee-social

[2] DIGI here refers to the criteria of a social media service as outlined in the *Online Safety Act 2021, https://www.legislation.gov.au/C2021A00076/latest/text*

not – this covers both harmful and illegal activity, enforcement actions that outline how platforms response to policy-violating content, and collaboration with civil society and government. While interventions will differ depending on the service, and are detailed in company's submissions to the inquiry, at a high-level platform-level approaches to online harms include:

2.2.1. Strict policies to prohibit and rapidly remove policy-violative or illegal content. These policies are regularly updated to ensure they reflect emerging patterns of abuse, in consultation with experts.

2.2.2. In many cases, the use of proactive technology detects the majority of problematic content proactively, and removes it before any user sees or reports it, or flags it for human review.

2.2.3. Reporting tools where content can be reported for action. Reports are reviewed by teams of human moderators with endeavours to be actioned as quickly as possible. Enforcement actions include the removal of content, and the suspension or removal of accounts that have instigated it.

2.2.4. Blocking tools where any user can be blocked from sending further unwanted messages, and provide tools to enable people to leave or hide group forums.

2.2.5. Safety-by design to mitigate online harms and in-build protections in the design of digital products.

2.2.6. Industry's policies and enforcement are complemented with a range of initiatives, partnerships and social programs.

2.3. DIGI believes that online safety must be advanced at the platform level through safety by design, a core principle that underlies all of DIGI's code development work including the Class 1 industry codes, along with other regulatory instruments under the OSA, industry standards for relevant electronic services and designated internet services (also for class 1A and class 1B material) and the BOSE.

2.4. DIGI wishes to emphasise that industry approaches to online harms will differ based on the services they provide, their users and their size. Not all services will experience the full range of potential online harms, and the way that different online harms present themselves on each service will differ, necessitating variations in approach. In this submission, we also include examples of work led by DIGI members as a complement to individual submissions made to this inquiry. These examples should not be taken as a comprehensive overview of all activities undertaken by members and we encourage the Committee to evaluate those for the relevant detail of company level responses.

2.5. Regulation provides an extremely important role in ensuring accountability for company-level responses to online harms, and ensuring industry-wide approaches that extend beyond mainstream services. DIGI supports strengthening the current online safety legislative framework 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. Appendix 1 also includes a list of as well as Australian Government regulation aimed at addressing the specific online harm, alongside the industry response. In addition, DIGI has also included a list of active processes underway that bear relevance to regulation of digital harms in Australia (Appendix 2, p.18).

2.6. It is important to remember that every online harm mirrors a distinct and complex social or economic policy issue that manifests online. While DIGI and its relevant members invest in prevention efforts through their partnerships, media literacy and training programs, the majority of platform-responses to online harms are 'downstream' in addressing the posting of that content online. DIGI believes in a holistic approach to online safety that also captures 'upstream' behaviours that can mitigate online harm; this is why we are a proponent of multi-stakeholder approaches in relation to online harms that continue to ensure strong accountability and responsibility on the part of online platforms, while also situating platform-level responses in a wider context that identities other actors and organisations that have important, additional roles to play.

3. Other technical measures to support age-appropriate experiences online

3.1. Young people especially must be protected from experiencing online harms. In this section, DIGI will discuss industry and measures related to online safety and young people. A discussion on children's data and privacy can be found subsequently in this submission (see section: Algorithms and online harms).

3.2. DIGI believes there are several industry approaches and legal safety nets (see Appendix 1) that are crucially important to protecting minors online. Of particular note are the industry approaches detailed in the Appendix that are of most relevance to protecting young people online, such as the rapid removals around Child Sexual Abuse Material (CSAM) cyberbullying, non-consensual intimate imagery, adult content and self-harm content. The Online Safety Act in particular creates a range of accountability mechanisms for industry's rapid response, including takedown schemes under the OSA relating to cyberbullying, CSAM and and a wide range of harmful content, where such content must be removed within 24 hours at the direction of the Office of the eSafety Commissioner. These takedown schemes provide an important safety net when such content is not rapidly removed as intended.

3.2.1. Mitigation of online harms are as important as the response. Product design features and tools are important to protect minors from age-inappropropriate content, such as adult content.

3.2.2. At the search engine level, Google's Safe Search filter[3] prevents search results containing or promoting nudity, sexually suggestive content, adult entertainment and other services from appearing within search results.

3.2.3. At the platform level, there are similar "safe search" settings that hide sensitive content and remove blocked and muted accounts. For example, YouTube age-restricts content that does not violate its Community Guidelines but may still not be appropriate for viewers under 18.[4]

3.2.4. One way to ensure parents can make decisions based on the evolving capabilities of children in their care is through increasing the prevalence and uptake of parental controls that give parents visibility about children's online activity, and opportunities to intervene;

[3] Google Search Help, *Filter explicit results using SafeSearch - Android - Google Search Help*, accessed at https://support.google.com/websearch/answer/510?hl=en&co=GENIE.Platform%3DAndroid. NB. Safe search will soon be turned on by default for all under 18 year old users in Australia.
[4] YouTube Help, Age-restricted content, https://support.google.com/youtube/answer/2802167?hl=en

DIGI members have extensive experience in developing and implementing such controls. As examples of some of the parental controls available:

3.2.4.1. At the service provider level, Apple[5] and Google[6] provide applications to enable family sharing and limitations on childrens' phones and tables, that include controlling their privacy settings, filtering access to content, screen time limits and other features designed to safeguard minors' privacy and experiences online.

3.2.4.2. At the app distribution level, restricted profiles can be established where more mature content can be filtered out of the app store.

3.2.4.3. On browsers, such as Chrome, parents can create restricted profiles for minors that allow parents to block and approve sites viewed, and Safe Search is on by default in such accounts.

3.2.4.4. At an app level, relevant DIGI members offer a variety of parental controls. For example, Snapchat's Family Centre lets parents see their teens' friends on Snapchat and who they have been communicating with over the last seven days, as well as a complete list of group members in group chats their teen is in that have been active over the last week.[7]On Instagram, parents can set a time limit for how long their teen can use Instagram each day or set scheduled breaks that limit their teen's use of Instagram during selected days and hours.[8]

3.2.4.5. There are also moves toward default privacy and safety settings for minors; for example, Instagram defaults teens into private accounts upon sign-up, and uses a number of safety measures for users in this category, including making it harder to adults to comment or interact with them, steps to inhibit inappropriate interactions with adults in private messaging, and preventing teens from seeing age-sensitive ads[9]. On Snapchat, contact settings for teens are set to friends and phone contacts only, and they can't be expanded to strangers.[10] TikTok doesn't allow direct messages to be sent to accounts held by under 16s. These are examples to illustrate the variety of product features and tools designed to support age-appropriate experiences on social media and should not be taken as a comprehensive list of all teen privacy and safety settings available..

3.3. DIGI has also observed that social media services are also introducing new teen-focused safety tools, such as Instagram's "Take A Break" feature that prompts young people who have been scrolling for a certain amount of time to take a break, suggests they set reminders to take more breaks in the future, and provides expert tips to reset.[11] Instagram also prompts teens to turn on 'Quiet Mode' which, once enabled, stops you from receiving any notifications, changes a profile's activity status to 'In quiet mode' and automatically sends an auto-reply to DMs.[12] TikTok has daily screen time set to 60 minutes which helps young people to be intentional about the time they spend on TikTok.Continued innovation

[5] Apple Support, *Use parental controls on your child's iPhone, iPad, and iPod touch*, accessed at: https://support.apple.com/en-us/HT201304
[6] Google, *Google Family*, accessed at: https://families.google.com/familylink/
[7] Snapchat 2024, What is Family Centre? https://help.snapchat.com/hc/en-gb/articles/7121384944788-What-is-Family-Centre
[8] Instagram 2024, parental supervision, https://help.instagram.com/309877544512275
[9] See the following links for more information: Youtube, *YouTube - More choices for families*, available at https://www.youtube.com/myfamily/; Meta, *New Teen Safety Features and 'Take a Break' on Instagram*, accessed at https://about.fb.com/news/2021/12/new-teen-safety-tools-on-instagram/; Twitter, Understanding and obtaining parental consent to use Twitter, available at https://help.twitter.com/en/using-twitter/parental-consent
[10] Snapchat 2024, Safeguards for Teens, https://parents.snapchat.com/en-GB/safeguards-for-teens
[11] Meta, "Raising the Standard for Protecting Teens and Supporting Parents Online" (7/12/2021), (Blog post by Adam Mosseri, Head of Instagram), accessed at https://about.fb.com/news/2021/12/new-teen-safety-tools-on-instagram/
[12] Instagram 2024, Giving People Control Over Their Time and What They See on Instagram, https://about.instagram.com/blog/announcements/new-ways-to-control-what-you-see-on-instagram

informed by evidence is important to ensure age appropriate experiences online.

3.4. DIGI notes that in addition to these measures, it is essential that children are supported to build digital literacy and capacities through education. We note and applaud the Government's investment in this area such as the Alannah & Madeline eSmart initiative.[13]These are lifelong skills that play an important role in how all Australians experience online environments.

4. Considering the best interests of the child

4.1. When considering online participation for young people in Australia, DIGI notes that the General comment No. 25 (2021) on children's rights in relation to the digital environment, as formally adopted by the United Nations Committee on the Rights of the Child (UNCRC) in February (2021),[14] provides best practice principles in assessing complex and interdependent factors that contribute to young people's experiences online.

4.2. Under Australia's current online safety regime, DIGI has supported the introduction of an additional expectation in subsection 6(2 A) of the BOSE concerning the best interest of the child, consistent with Article 3 of the UNCRC. We also understand that 'the best interests of the child' covers the full range of a child's rights, including a right to safety, to privacy, to free expression, to access information, to autonomy, and to a range of other critical rights.

4.3. The General Comment states that 'meaningful access to digital technologies can support children to realise the full range of their civil, political, cultural, economic and social rights. However, if digital inclusion is not achieved, existing inequalities are likely to increase, and new ones may arise.'[15]This highlights the need to consider the right settings to not only prevent harm, but to support young people to access the benefits and opportunities of digital technologies. We note that the Class 1 Codes developed by industry under the OSA impose obligations on providers subject to the Codes to consider 'the rights and best interests of children' more generally when implementing the relevant measures. This ensures that providers give consideration to the full range of childrens' rights online as set out in General comment No 25.

4.4. DIGI encourages the committee to consider what policy settings and initiatives can support healthy and meaningful experiences for young people online. Australian children's safety and research organisations are actively examining this issue. For example, research by University of Western Sydney's Young & Resilient Centre, funded by the Office of the eSafety Commissioner notes the importance of fostering digital capacities for online safety and to help deliver positive experiences for young people online, stating 'learning how to communicate effectively and respectfully is crucial for managing a range of online and offline safety concerns, as well as cultivating (digital) citizenship and an ethics of (online) engagement'.[16]We recommend that the committee considers the crucial role of education and capacity building in providing young people with the tools to thrive in the digital age.

[13]Alannah & Madeline 2024, The importance of teaching media literacy - A guide for Australian teachers, https://www.alannahandmadeline.org.au/resources/the-importance-of-teaching-media-literacy-a-guide-for-australian-teachers
[14] Ibid
[15]https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation#:~:text=In%20this%20general%20comment%2C%20the,the%20Optional%20Protocols%20thereto%20in
[16] https://www.westernsydney.edu.au/__data/assets/pdf_file/0004/1976503/Reimagining_Online_Safety_Education.pdf

4.5. DIGI supports the recent comments of the eSafety Commissioner during Senate Estimate hearings, highlighting the complexities in the topic of age bans, including the challenges in reaching a robust evidence base to support a specific age limit.[17] These comments raised, among other issues, that eSafety research has shown 'young people, particularly those who are from disadvantaged backgrounds or Indigenous or who identify with a disability or who are LGBTQI+ tell us that they feel like they are more themselves online than they can be in real life.'[18]This highlights that there are contextual factors that can contribute to young people's experiences both online and offline and that, for some demographics, social media can be an important channel for community and social inclusion. We also note that these demographics can be at heightened risk of online abuse and the necessity to implement robust proactive and reactive safety measures to reduce their exposure to potentially harmful material online.[19]We encourage the committee to consider how age-related bans might affect an individual minor differently based on a number of contextual factors, including race, ethnicity, and sexuality, and of unintended consequences that might result from excluding young people from online spaces.

4.6. DIGI further notes that there are a variety of perspectives in research findings, and young people and parents' experiences and support further engagement with experts, young people, parents and industry in examining what age-appropriate experiences might look like in a variety of settings. We support the need to meaningfully consult with young people on any digital policy relevant to their online participation and use of digital technologies, a position that has been raised by many leading children's safety organisations such as PROJECT ROCKIT[20] and Alannah & Madeline Foundation[21] and echoed by the delegate representing Children and Young People with Disability Australia at the 17th Session of the Conference of States Parties (COP17) to the United Nations (UN) Convention on the Rights of Persons with Disabilities in June this year.[22] It's essential that any proposals are evidence informed to lead to meaningful safety outcomes. DIGI supports further research and consultation, including with young people, is pursued to ensure a robust evidence-informed approach to any proposal related to age-appropriate online participation.

4.7. To this end, DIGI acknowledges that that the Government is undertaking a pilot trial of age assurance technology and that this is an appropriate avenue to consider the privacy and technical challenges related to the use of age verification technologies, including its privacy risks related to identity verification, through the provision of drivers' licences, passports or other Government issued identification documents and/or biometric data. DIGI supports the Government's intention to further scope the efficacy and potential risks

[17]30 May 2024, Senate Estimates, P.58
https://www.aph.gov.au/-/media/Estimates/ec/bud2425/30_MayEnvironment_and_Communications_Legislation_Committee_2024_05_30.pdf?la=en&hash=4257C26084262A905D15E0899EC3D465C02093EB
[18] Ibid.
[19] eSafety 2021, Protecting LGBTIQ+ voices online,
https://www.esafety.gov.au/sites/default/files/2021-08/LGBTQI%2B%20cyber%20abuse%20resource%20development%20-%20Report.pdf
[20]PROJECT ROCKIT 2024,
https://www.linkedin.com/posts/lucylockit_is-social-media-good-or-bad-the-most-activity-7208791007927128064-Gnel?utm_source=share&utm_medium=member_desktop
[21]Alannah & Madeline Foundation 2024, Should we raise the age for social media? Why that's the wrong question,
https://www.alannahandmadeline.org.au/news/should-we-raise-the-age-for-social-media-why-thats-the-wrong-question
[22]17th Session Of The Conference Of States Parties To The CRPD (COSP17),
https://social.desa.un.org/issues/disability/cosp/17th-session-of-the-conference-of-states-parties-to-the-crpd-cosp17

of age assurance technologies, including the potential for unintended consequences.

4.8. We urge the committee to consider the full range of children's rights in relation to the digital environment and be cautious of the unintended consequences, such as limiting the capacity for young people to access information, connect with communities, and develop digital citizenship and other capacity building experiences.

5. Young people's digital habits and mental health

5.1. DIGI recognises the Committee and community concerns in relation to increased rates of recorded psychological distress in young people, as well as interest in youth mental health and social media specifically, and agrees that these are extremely important issues. Many of our members conduct research and partnerships with experts in this area to inform their work, and these are detailed in their own submissions to the inquiry.

5.2. We consider that current national conversations about young people's use of social media are important in raising awareness of the active role that the technology industry, parents and young people themselves, play in ensuring the healthy use of technology and positive experiences online. Government recommendations made in relation to youth mental health and social media should be evidence-informed and situated within a whole-of-Government approach, and we acknowledge the significant work underway by the National Mental Health Commission, including in examining digital wellbeing for Australian young people.[23] As the Commission notes in its recent Discussion Paper, 'research into the ways young people engage with digital technologies and the positive or negative impacts this can have is still in early stages'.[24]

5.3. Our understanding is that research to date has not established a direct causal link between social media use and youth mental health issues in Australia or globally[25], and that the existing research indicates a level of complexity in the interaction with other factors, such as the. social contexts and familial conditions in which children and young people live.[26] We urge caution at attributing any one factor as the cause of mental health issues.

5.4. DIGI is very supportive of further Australian research that aims to understand the link between digital technology use and the type of interventions that would be effective in addressing that link. Research into the mental health impacts of social media should also examine different cohorts of young people in Australia, as it is important that young people's online experiences are not studied in isolation from their lives in general. A UNICEF Discussion Paper summarised the necessity for such an approach, stating:

> *Researchers need to consider children's life contexts and socio-demographics to the greatest extent possible. More control variables need to be included in quantitative studies to ensure that variables that have known effects on child*

[23]National Mental Health Commission 2024, Discussion paper: Digital technologies and youth mental health, https://www.mentalhealthcommission.gov.au/sites/default/files/2024-05/discussion-paper---digital-technologies-and-youth-mental-health.pdf
[24] Ibid.
[25]2023, Andrew K. Przybylski, Oxford Internet Institute, Association for Psychological Science, Global Well-Being and Mental Health in the Internet Age, https://journals.sagepub.com/doi/10.1177/21677026231207791?icid=int.sj-full-text.citing-articles.4
[26] Swist, T., Collin, P., McCormack, J., & Third, A. (2015), *Social Media and the Wellbeing of Children and Young People: A Literature Review*, accessed at https://researchdirect.westernsydney.edu.au/islandora/object/uws:36407

well-being outcomes are not excluded. Children's online experiences cannot be studied in isolation from their lives in general.[27]

5.5. This research can underpin further Government, industry and civil society work in this area to ensure work is targeted and effective in improving youth mental health, and DIGI believes that the digital industry would be open to collaborating on this effort.

5.6. The default age for social media use being 13 has evolved as a norm because of a US law called the Children's Online Privacy & Protection Act (COPPA). DIGI sees value in conversations occurring at a national, industry and household level as to whether 13 is the appropriate age for individual children. DIGI is not in a position to comment on the appropriate age across the sector, especially given that 'social media services', as defined in Australia under the Online Safety Act, encompass a wide diversity of services. The Act has a broad definition of social media services as those with *"the sole or primary purpose of the service is to enable online social interaction between 2 or more end-users; (ii) the service allows end-users to link to, or interact with, some or all of the other end-users; (iii) the service allows end-users to post material on the service"*[28]

5.7. We would encourage a balanced and holistic view that draws upon expert perspectives, including the eSafety Commissioner and the National Mental Health Commission, ensuring that findings are shared with the digital industry to consider. Policy proposals here need to thoughtfully engage with experts, young people, parents and industry, and we welcome the opportunity to participate in those discussions.

5.8. In relation to mental health specifically, DIGI also notes that social media is a channel that young people use to access mental health support and services. A recent report from the University of Chicago found that 'young people are proactively taking charge of their own mental health by seeking information online via social media and mental health apps', including as a way to overcome barriers in accessing mental health services in offline contexts.[29]DIGI notes that a broad ban of young people from online spaces, including social media, may have an unintended consequence on increasing barriers to accessing mental health resources and services that are available through these channels.

5.9. None of that is to imply that research is required as precursor to action; Relevant DIGI members conduct significant work related to mental health. They have policies and enforcement mechanisms that prohibit pro-suicide, self-harm and pro-eating disorder and other harmful content. They have interception processes that routinely direct users identified as at risk to local support resources; for example, certain content searches and flags will direct users to obtain help through expert services such as Lifeline, tools, programs, outreach. For example, Snapchat's in-app mental health portal, Here for You, provides resources from expert organisations, including NGOs based in Australia, for getting help or supporting a friend in crisis.[30]On Instagram, if someone searches for terms related to disordered eating, they will see expert-based research first (including contacts for local eating disorders hotlines, such as the Butterfly Foundation in Australia)

[27] https://www.unicef.org/innocenti/media/8181/file/UNICEF-Innocenti-Time-Using-Digital-Tech-Impact-on-Wellbeing-2017.pdf
[28] *Online Safety Act 2021*, see Section 5.
[29]2024, Getting Help Online: How Young People Find, Evaluate, and Use Mental Health Apps, Online Therapy, and Behavioral Health Information,
https://www.commonsensemedia.org/sites/default/files/research/report/2024-getting-help-online-hopelab-report_final-release-for-web.pdf
[30] Snapchat 2024, Wellbeing Features,
https://help.snapchat.com/hc/en-us/articles/7012398974612-Wellbeing-Features-on-Snapchat

before seeing the search results.[31] They also invest in partnerships to deliver youth resilience-building programs, in areas such as cyberbullying and digital literacy and anti-bullying. These initiatives should be detailed in member companies' submissions to the inquiry.

6. Algorithms and protections for children's data and privacy

6.1. DIGI will discuss algorithms more broadly in section 9 of this submission but, we also acknowledge the particular community concerns regarding the role that algorithms and recommender systems have in influencing the content accessed by children. DIGI agrees this needs to be considered in the context of a robust online safety framework through principles-based, appropriate and enforceable regulation.

6.2. In addition, DIGI agrees there needs to be widespread protections for children's privacy and their data. DIGI sees a key opportunity to standardise these protections in the current Privacy Act Review and has long supported specific privacy protections for minors be expanded upon within the Privacy Act, in order to provide minors, or their guardians, confidence that a baseline standard of privacy exists no matter which online service they are using.

6.3. DIGI is supportive of the introduction of a Children's Online Privacy Code that considers effective practices from other markets and reflects the risks and benefits of different processing purposes to children. DIGI agrees that minors require additional privacy protections under the revised Act, and we support a code that draws on the UK 's Age Appropriate Design Code and other comparable models, which requires services to have regard to the best interests of the child.

6.4. As noted, DIGI has supported the introduction of an additional expectation in subsection 6(2 A) of the BOSE concerning the best interest of the child, consistent with Article 3 of the UN Convention on the Rights of the Child (UNCRC). Our understanding of 'the best interests of the child' is that it concerns whatever is best for that individual child. In some instances it may be best to restrict a particular minor from a service, however in other instances it might be in the best interests of the child to have access to a digital service.

6.5. Parents or guardians can make judgements on what is in the best interests of the child, not APP entities, because service providers rightly do not have that knowledge. DIGI considers that more research guidance would be valuable in considering how companies across the digital industry apply the best interests of the principle to children in a general sense in the digital world; at a minimum, this should include robust response to the online harms enumerated in Table 1. In addition, it should consider how safety-by-design can support the best interest of the child.

7. Multi-stakeholder approaches to address risk

7.1. DIGI believes a holistic approach to online safety that includes prevention and education measures is essential to mitigate online harm. We are a proponent of multi-stakeholder approaches in relation to online harms that continue to ensure strong accountability and responsibility on the part of online platforms, while also situating platform-level

[31] Meta 2024, How were supporting people affected by eating disorders, https://about.fb.com/news/2021/02/supporting-people-affected-by-eating-disorders-and-negative-body-image/

responses in a wider context that identifies other actors and organisations that have important, additional roles to play. Online harms are multi-faceted social problems and must also consider the relationship of offline and online behaviours to address and mitigate the risk of harm.

7.2. DIGI and its relevant members also invest in prevention efforts (e.g. through partnerships, media literacy and training programs) and DIGI's relevant members have long standing research and community partnerships. For example, DIGI has worked on the prevention of online harms through our DIGI Engage youth summits. Over multiple years, DIGI partnered with the Australian Government – through the Department of Home Affairs, Multicultural NSW and the Attorney General's Department – around our shared goals in relation to countering violent extremism. Previous DIGI Engage youth summits have upskilled hundreds of young people about the root causes of societal polarisation, hate speech and extremism, and has built their capability online and offline to counter them. In 2024, we are proud to be partnering with Multicultural NSW to relaunch the concept as COMPACT X DIGI Engage. DIGI is partnering with the NSW Government's COMPACT Program, which is a best practice community resilience-building approach to addressing the same issues. This partnership focuses on digital community building, co-design, upskilling, and knowledge sharing for both practitioners working to counter hate in their communities, and the young people with whom they work.

8. DIGI's work to combat disinformation and harmful misinformation

8.1. The Australian Code of Practice on Disinformation and Misinformation

8.2. In this section, DIGI will outline our work to combat digital harms related to disinformation and misinformation. DIGI and its members share a strong commitment to ensuring the transparency and integrity of Australian democratic political processes, and institutions recognising that as important actors in the information ecosystem, they have a critical role and responsibility in reducing the spread of disinformation and misinformation online. The ACPDM was launched in February 2021 in response to government policy as set out in *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry.*[2]

8.3. DIGI developed this code with assistance from the University of Technology Sydney's Centre for Media Transition[32] and First Draft[33], a global organisation that specialised in helping societies overcome false and misleading information.

8.4. The ACPDM adopts an outcomes based approach that aims to incentivise signatories to be more transparent and accountable for their response to harms caused by disinformation and misinformation. To date, the code has been adopted by Apple, Adobe, Google, Meta, Microsoft, Redbubble, TikTok, Twitch and most recently Legitimate. These companies have all committed to implement safeguards to protect Australians against online disinformation and misinformation. Mandatory code commitments include:

[32]UTS 2024, Centre for Media Transition, https://www.uts.edu.au/research/centre-media-transition
[33]First Draft 2024, About First Draft, https://firstdraftnews.org/about/

Publishing & implementing policies on misinformation and disinformation, providing users with a way to report content against those policies and implementing a range of scalable measures that reduce its spread & visibility (Mandatory commitment #1). Every signatory has agreed to prepare annual transparency reports about those efforts to improve understanding of both the management and scale of mis- and disinformation in Australia (Mandatory commitment #7).

8.5. Additionally, there are a series of opt-in commitments that platforms adopt if relevant to their business model: (Commitment #2) Addressing disinformation in paid content; (#3) addressing fake bots and accounts; (#4) transparency about source of content in news and factual information (e.g. promotion of media literacy, partnerships with fact-checkers) and (#5) political advertising; and (#6) partnering with universities/researchers to improve understanding of mis and disinformation. DIGI remains supportive of ACMA oversight for the ACPDM to be formalised in legislation.

8.6. DIGI considers that political communication is fundamental to the proper functioning of Australia's democratic processes. The ACPDM helps uphold this important function by addressing content that could fall under the scope of foreign interference in democratic processes. Under the code, 'disinformation' is defined as 'digital content that is verifiably false or misleading or deceptive that is propagated amongst users of digital platforms via inauthentic behaviours, the dissemination of which is reasonably likely to cause harm' and the definition of 'harm' includes 'democratic political and policy making processes such as voter fraud, voter interference, voting misinformation'. These definitions of content the code applies to are a critical step in identifying and tackling the potential for foreign interference in online discourse.

8.7. In addition, there is the challenge of differing community expectations with regard to certain online harms which are compounded when applying policies at a large scale. For example, one of the biggest challenges that DIGI encountered in developing the ACPDM was there is no consensus as to what constitutes misinformation and disinformation – this is an area where academics, regulators, MPs and media all disagree. In light of differing views, DIGI focused its definition of disinformation and misinformation on that which crosses a threshold of harm. We recognise that some people will consider this approach as impinging on freedom of expression, while others will believe that the definition does not go far enough to capture everything perceived as misinformation. Any regulation to address mis- and disinformation will need to grapple with this challenge.

8.8. The ACPDM has been updated and strengthened over time through periodic review and collaborative analysis on its implementation. The ACPDM framework was designed to respond rapidly to advances in technology and the threat landscape, as well as recognise that the issue of disinformation and harmful misinformation manifests differently depending on the context of the product, service, or information environments. DIGI is committed to driving improvements in the management of mis- and disinformation in Australia and further strengthening the industry code and we support a regulatory backstop to the code through increased powers for the ACMA.

8.9. In October 2021, DIGI put in place governance arrangements to strengthen the ACPDM and its effectiveness. DIGI appointed an independent complaints committee to resolve complaints about possible breaches by signatories of their code commitments, and created a portal on DIGI's website for the public to raise such complaints. We appointed an independent reviewer to fact check and attest all signatories' transparency reports prior to publication, who also developed best practice reporting guidelines to drive

improvements and consistency in 2022 transparency reports.

8.10. Complaints raised under the code by the public can be assessed by an independent complaints committee, with positions held by Anne Kruger, Victoria Rubensohn and Christopher Zinn[34], and the scope and attestation of the signatories' transparency reports is undertaken by an independent reviewer, Hal Crawford.[35] These independent experts have been appointed because of their extensive experience and subject matter knowledge. DIGI's role in the complaints process is administrative and it cannot vote on committee decisions.

8.11. In its June 2021 *Report to government on the adequacy of digital platforms disinformation and news quality measures* (ACMA Report to Government), released in March 2022, the ACMA reviewed the ACPDM, finding that 'the code objectives and principles meet the government objective of striking a balance between encouraging platform interventions and protecting freedom of expression, privacy and other rights.'[36] DIGI has supported, in principle, the ACMA's recommendations to the Government to have greater oversight of the code and misinformation more broadly, as a complement to existing, robust self regulation measures.

8.12. In its report to Government, the ACMA advanced five key recommendations to strengthen their oversight of the code and their work on misinformation and disinformation. In June 2022, DIGI commenced a review of the code, which included close consideration of these recommendations, as well as submissions received as part of a public consultation process.[37]

8.13. Some of the key changes DIGI made as part of the review of the code include:

8.13.1. Encouraging greater participation in the code by smaller digital platforms, including by modifying the transparency reporting requirements for services with less than one million active monthly users in Australia. This acknowledged the likelihood for misleading content to proliferate elsewhere online as mainstream platforms strengthened approaches to tackling misleading content.

8.13.2. Updated definition of 'harm' in relation to mis and disinformation, addressing stakeholders' concerns that the threshold of 'serious and imminent' threat of harm was too high; the new threshold is 'serious and credible' threat of harm. For the purposes of this Inquiry, we note that the definition of harm has always encompassed voter interference.

8.13.3. Additional commitments reflecting updates to the strengthened EU Code of Practice in relation to recommender systems, and deterring advertisers from repeatedly placing digital advertisements that propagate mis- and disinformation. There are also updates to further clarify that both sponsored content and paid for advertising are in scope of relevant commitments on demonetising mis- and disinformation.

8.13.4. Retaining the pre-existing exclusion of professional news content from being treated as misinformation under the code, and the pre-existing obligation for signatories to address this content when it is being disseminated as disinformation. The review concluded that the ACMA and the professional news

[34] DIGI 2024, https://digi.org.au/disinformation-%20code/governance/
[35] Ibid.
[36] ACMA 2022, *Report to government on the adequacy of digital platforms' disinformation and news quality measures*, https://www.acma.gov.au/report-government-adequacy-digital-platforms-disinformation-and-news-quality-measures
[37] DIGI (2022), *Code Review,* https://digi.org.au/disinformation-code/code-review/

media are best placed to address misinformation concerns within their self regulatory and co-regulatory codes.

8.13.5. Requiring greater transparency around the specific products and services that are within scope of the signatories' code commitments, through updates to the code, transparency reporting requirements and the DIGI website. [38]

8.14. Signatories remain committed to continuously improving transparency reporting under the Australian Code of Practice on Disinformation and Misinformation, consulting with the ACMA on the transparency reporting process and the development of suitable metrics to assess signatories' performance against the code outcomes. For example, in 2024, the best practice guidelines were updated to more explicitly cover measures taken to combat disinformation and misinformation generated by artificial intelligence, as well as to address significant policy changes related to disinformation and misinformation that have taken place since the last reporting period.[39] These updates are the latest set in a series of improvements driven by DIGI and code signatories since the code was introduced in February 2021.

8.15. DIGI recognises that more progress can be made. Digital harms require continued and increasing investment through teams and technology, and research into emerging patterns and evolving community standards to inform continual iteration and improvement.

8.16. To illustrate the need for multi-stakeholder approaches, continuing with the example of misinformation, there is a complex interplay between traditional media and digital platforms on this challenge, as well as other stakeholders. Research by UTS and First Draft shows the hashtag "arson emergency" was propagated by 300 inauthentic social media accounts as disinformation about the cause of Australia's devastating summer of bushfires, but the claim was also published by news outlets.[40] More recent examples include the repetition of defamatory claims on traditional media in the aftermath of the Bondi Junction attacks. We also support the ACMA having greater powers as a regulator in relation to misinformation; a comprehensive approach that includes restrictions and oversight on misinformation across different industries that the ACMA oversees will ensure an ecosystem approach to misinformation. A wide policy lens in our approach – that examines platforms alongside all relevant actors in the ecosystem – to all online harms will move us further forward to the outcomes we seek, and is important as we consider how policy approaches can be strengthened.

9. Algorithms and potential related risks

9.1. Noting the committee's interests in the extent to which algorithms used by social media platforms permit, increase or reduce online harms to Australians, this section offers an exploration of the usage of algorithms, as well as recommendations for how the Australian Government might consider mitigating harm. DIGI isn't in a position to comment on the specific function of recommender systems at a platform level but

[38] Ibid.
[39] DIGI 2024, *Annual Tech Sector Transparency Reports Give Insight Into The Online Misinformation Fight,* https://digi.org.au/in-the-media/2802/
[40] UTS, *Discussion Paper on an Australian Voluntary Code of Practice for Disinformation,* Prepared for DIGI by UTS Centre for Media Transition, https://digi.org.au/wp-content/uploads/2021/02/Discussion-Paper-ACPDM-FINAL-PDF-Updated-Feb-2021.pdf, p. 16.

considers the topic in the broader context of online safety.

9.2. DIGI believes that consumers should be provided with information about why they are seeing particular content. On social media, we also support tools that help people exercise choice and control over their exposure to certain types of content (for example, sensitive content). DIGI members also have tools in place to help Australians have more control and choice in regards to what they see. For example, Meta offers a range of tools and transparency measures across its services, including tools to manage how content ranks in your Facebook Feed.[41] On Instagram, there are sensitive content controls that allow you to decide how much sensitive content shows up in *Explore* and when you select "Not interested" on a post seen in *Explore*, it will also aim to avoid showing you this kind of content going forward in other places where recommendations are made like Reels and Search.[42]Meta also provides information to users to help them better understand how its ranking algorithms and AI-powered products work and when they are engaging with AI-generated content. For example, they have released more than 22 AI System Cards that explain how the AI systems in their products work.[43]It also made information detailing the ranking process on Instagram from start to finish publicly available via a blog.[44] On YouTube, you can turn off your watch history or search history when you don't want your watches or searches to influence future recommendations and search results.[45] TikTok also has a variety of content controls available to help Australians choose what they wish to see on their TikTok feed.[46]

9.3. In addition, DIGI agrees with the need for risk-based frameworks to prevent and address issues related to the use of Artificial Intelligence (AI) and Automated Decision Making (ADM), in a wide range of areas including preventing discrimination. DIGI also notes that proposals relating to algorithmic transparency have been agreed in principle by the Government in reforming The Privacy Act 1988. This year, the Government also announced development of a voluntary AI Safety Standard options for voluntary labelling and watermarking of AI-generated materials.

9.4. On relevant large digital platforms, AI and algorithms play an important role as a sorting mechanism for the millions of terabytes of information online, enabling people to readily obtain relevant content and information. AI is also used to safeguard the safety and security of Internet users, and to address harmful content. Such technology is having a positive effect; In Q4 of 2023, approximately 99.6% of the comments removed from YouTube were detected by automatic flagging. Automated flagging also allows videos that violate community guidelines to be removed before they are widely viewed. In the same quarter, 51.51% of the videos flagged by automation were removed before they received a single view.[47] Similarly, Meta uses artificial intelligence to proactively detect harmful content before it is seen by users. For example, in Q1 2024, Meta proactively detected 14.4 million pieces of child sexual exploitation content on Facebook, 94.3

[41]Facebook 2024, Facebook Help, https://www.facebook.com/help/543114717778091
[42] Instagram 2024, Giving People Control Over Their Time and What They See on Instagram, https://about.instagram.com/blog/announcements/new-ways-to-control-what-you-see-on-instagram
[43] Meta 2024, Our approach to explaining ranking, https://transparency.meta.com/features/explaining-ranking/
[44]Instagram 2021, Shedding more light on how Instagram works, https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works
[45]YouTube Help 2024, Manage your recommendations & search results, https://support.google.com/youtube/answer/6342839?hl=en&co=GENIE.Platform%3DAndroid
[46]TikTok2024, Content Controls, https://www.tiktok.com/safety/en/content-controls/
[47] YouTube (2023), YouTube Community Guidelines enforcement, accessed at https://transparencyreport.google.com/youtube-policy/removals

percent of which was found and actioned before a user reported it.[48] Microsoft's integration of GPT4 enabled content moderation solutions for Microsoft Start service helps proactively block violative content.[49] Proactive detection is also being used extensively to identify and prevent scams targeting Australian consumers. For example, Gmail proactively blocks more than 99.9% of spam, phishing, and malware before it reaches users.[50]

9.5. The use of algorithms to promote online safety is consistent with the Government's expectations of industry. For example, the BOSE determination, which came into effect with the Online Safety Act (OSA) on January 23, 2022, identifies the detection of material and activity as a reasonable step service providers can take to ensure end users are safe.[51] The Office of the eSafety Commissioner's Safety by Design principles include "Using scanning and filtering technology to ensure user safety is upheld on the site and users are not exposed to inappropriate or sensitive content."[52]The registered *Social Media Services Online Safety Code (Class 1A and Class 1B Material)*, developed under the OSA, also includes use of artificial intelligence and machine learning as safety measures related to the detection and removal of child exploitation and pro-terror material.[53]

9.6. It is important to note that algorithms also do not operate in isolation from human intervention; in relation to content removal, it is often the case that AI surfaces problematic content for a human moderator to review for context and accuracy, and to guide the most effective decision. AI plays an important role in scanning content at a scale that humans could never achieve, at a speed which was previously not possible. It forms a key part of how online safety challenges are addressed at a large scale.

10. DIGI's work to combat scams and improve consumer protection

10.1. As previously stated, DIGI sees a key goal of our organisation as helping convene major digital industry players, in close collaboration with Government and community stakeholders, around shared goals, including maximising the benefits and opportunities of tech for all Australians, delivering strong consumer protections online, and supporting a thriving digital enabled-economy.

10.2. According to Australians' reports to Scamwatch, released in the recent National Anti-Scams Centre (NASC) Quarterly Update, text message remains the most popular

[48]Meta (2023), "Community Standards Enforcement Report Q1 2024", accessed at https://transparency.fb.com/data/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/
[49] Microsoft 2024, ACPDM annual transparency report, https://digi.org.au/wp-content/uploads/2024/05/Microsoft-LinkedIn-Australian-Disinformation-Misinformation-Code-2024.pdf
[50] https://workspace.google.com/blog/identity-and-security/how-gmail-helps-users-avoid-email-scams
[51] Department of Infrastructure, Transport, Regional Development and Communications, *Draft Online Safety (Basic Online Safety Expectations) Determination 2021*, accessed at https://www.infrastructure.gov.au/have-your-say/draft-online-safety-basic-online-safety-expectations-determination-2021-consultation
[52] Office of the eSafety Commissioner, "Safety by Design | Principles and background", accessed at https://www.esafety.gov.au/industry/safety-by-design/principles-and-background
[53] Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material (2023) https://onlinesafety.org.au/wp-content/uploads/2023/06/230616_1_SMS-Schedule_REGISTERED-160623.pdf

method of choice for scammers (34%), followed by phone call (27%), email (22%), internet sources such as websites (6%), followed by social media and online forums such as trading sites (taken together 6%).[54] We recognise that cross-sectoral collaboration is essential to target and combat financial fraud, particularly as scammers are adept at moving from service to service to avoid detection and disruption efforts.

10.3. DIGI's relevant members have long standing, multi-pronged anti-scam efforts that include enforced restrictions to rapidly combat scams. This could include through proactive detection, in-product reporting and customer service, as well as digital literacy efforts to reduce Australians' susceptibility to scams. They also invest significantly in cyber security that protects consumers against a range of harms, including scams. Their restrictions on scams also include spam, fraud and other deceptive conduct – including phishing, impersonation and misrepresentation – on organic content as well as paid content and advertising. Many of them work closely with other companies and governments, including with the ACCC's Scamwatch program and the National Anti-Scam Centre, to both identify and act on trends in scams and criminal behaviour. DIGI includes a short overview of this work on its website.[55]

10.4. In 2024, DIGI prepared a new publicly available 'how to report scams on digital platforms' guide for Australian consumers.[56]The goal of this document is to create a quick reference point that helps guide consumers to the right place for each of our relevant members. We shared this in the DIGI newsletter to socialise it with relevant stakeholders and it is publicly available on the DIGI website.

10.5. DIGI supports the work of the NASC and is proud to be represented on its Advisory Board, its Data Integration and Technology Working Group, and its Communication and Awareness Working Group. We are supportive of the 'ecosystem' approach the NASC takes to foster close collaboration between industry and government. As scams can span multiple services, regulatory approaches should be holistic in involving a range of relevant industries across the private sector as well as consumer bodies, regulators and law enforcement. The establishment of the NASC has fostered new levels of cross-sectoral collaboration and intelligence sharing, as well as strong ties between industry, government and community organisation, an approach that has seen promising signs as annual and quarterly losses to scams are both recorded on a declining trend.[57]

10.6. DIGI shares the Government's goal in seeking to lift the bar to ensure robust and effective approaches to scams in relevant industries. Financial criminals manipulate multiple sectors simultaneously to defraud Australians, and we welcome the funding announced in the recent Federal Budget to ensure improved responses cross-sectorally to complement the important work being undertaken by the National Anti-Scam Centre. DIGI supports mandatory industry codes that are evidence-based, appropriately targeted and proportionate in key sectors to help lift accountability in relation to scams across key sectors, as well as the ability of sectors to work together to combat scams. We will continue to work with the Government on clear digital industry obligations that will drive better outcomes for Australians. DIGI has convened an industry working group to consider what implementable industry level obligations might look like in the diverse

[54]ACCC 2024, NASC Quarterly Update January - March 2024, https://www.accc.gov.au/system/files/NASC-Quarterly-update-Q3-2024.pdf
[55] DIGI 2024, Scams and consumer protection, https://digi.org.au/scams/
[56] Ibid.
[57]ACCC 2024, NASC Quarterly Update January - March 2024, https://www.accc.gov.au/system/files/NASC-Quarterly-update-Q3-2024.pdf

digital industry, as a complement to, and in the context of the Government's legislative agenda in relation to scams.

11. Conclusion

11.1. DIGI and its members recognise the immense importance of addressing online harm and the necessity for industry, experts, community organisations, and Government to work together in doing so. DIGI sees itself as a key Government partner in this endeavour, through the work outlined above, and our ongoing engagement on the development of many pieces of regulation and legislation where we advocate for approaches that are effective in their goals and can practically be implemented by industry.

11.2. DIGI welcomes this inquiry as a way to have a deeper conversation about online safety, and we hope that this submission assists in advancing a shared understanding of the current landscape, themes and potential gaps. Online harms are multi-faceted social problems that cannot be fixed with technical and legal safeguards alone; this is why we are a proponent of multi-stakeholder approaches in relation to online harms that continue to ensure strong accountability and responsibility on the part of online platforms, while also situating platform-level responses in a wider context. DIGI looks forward to continuing to collaborate with a range of stakeholders on our shared goals in combating online harms, including through many of the active processes outlined below.

Appendix 1 - Online harms and regulatory responses

Table 1: Overview of online harms, platform & regulatory responses		
Online harm description	**Trends in platform responses**	**Australian regulatory responses**
1. Cyberbullying material directed at an Australian child	1.1. All relevant DIGI members have strict policies to prohibit and rapidly remove the cyberbullying of Australian children and minors. These policies are regularly updated to ensure they reflect emerging patterns of abuse, in consultation with experts. 1.2. They provide reporting tools where content can be reported for cyberbullying. Such messages are reviewed by teams of human moderators, and addressed as quickly as possible. Enforcement actions include the removal of cyberbullying content, and the suspension or removal of accounts that have instigated it. 1.3. This enforcement infrastructure is often complemented with proactive technology detection that detects problematic content and flags it for human review. 1.4. Relevant members provide blocking tools where any user can be blocked from sending further unwanted messages, and provide tools to enable people to leave or hide group forums. 1.5. Industry's policies and enforcement are complemented with a range of initiatives, partnerships and social programs aimed at providing minors with wider support from professionals, parents and teachers in relation to cyberbullying.	1.6. Under the Online Safety Act 2021, Australian minors who are the target of cyberbullying material, and those representing them, can complain to the Office of the eSafety Commissioner (the Commissioner). The Commissioner can direct a request for removal to the social media service, and the service must remove the content within 24 hours. 1.7. The Basic Online Safety Expectations (BOSE) apply to all social media services, messaging services and websites. A core expectation of the BOSE is that a provider of a service must take reasonable steps to minimise the extent to which cyberbullying material targeted at an Australian child or adult is available, and to make reports about the provider's related activities available to the Commissioner. 1.8. The OSA children's cyber bullying scheme enables the Commissioner to issue end-user notices that require a person who posts cyberbullying material to remove the material, refrain from posting any cyberbullying material targeting the child, and/or apologise to the child for posting the material. 1.9. Section 474.17(1) of the Criminal Code 1995 (Cth) creates an offence of using a carriage service to menace, harass or offend another person.

2. Child sexual abuse material (CSAM)	2.1. DIGI members have zero tolerance for CSAM. They have strict policies against child exploitation and the sexualisation of children. These policies are enforced through proactive detection technology and human review teams who undergo extensive training on the appropriate protocols for the handling of CSAM material. The vast majority of this material is removed by proactive detection technology without users ever having seen it. 2.2. When CSAM is detected it is removed and reported, DIGI members report to the National Center for Missing & Exploited Children (NCMEC) in the United States which refers cases to law enforcement all around the world, including in Australia. They also directly cooperate with Australian law enforcement operations. 2.3. Relevant DIGI members are active in several coalitions, such as the Technology Coalition, the ICT Coalition, the WeProtect Global Alliance, and INHOPE and the Fair Play Alliance, that bring companies and NGOs together to develop solutions that disrupt the exchange of child sexual abuse materials online and prevent the sexual exploitation of children. 2.4. Relevant DIGI members deploy industry-developed and licensed technological tools such as Photo DNA (developed by Microsoft to identify known CSAM in still images) and CSAI Match (developed by YouTube to detect known video-based CSAM).	2.5. The OSA includes a removal scheme for child sexual exploitation material. The Commissioner can direct a request for removal to the social media service, and the service must remove the content within 24 hours. 2.6. Furthermore, the OSA Codes (detailed earlier in this submission, relate to “Class 1” and “Class 2” materials under Australia’s classification code. The list of Class 1 materials includes CSAM. DIGI co-led the development of mandatory codes required under the Online Safety Act, that Autralians with new enforceable protections against child sexual exploitation material, pro-terror content and other extremely harmful materials (Class 1 codes). Six of these codes are now in force, creating new online harms legal obligations for social media services, app distribution services, search engines, equipment providers and related services, hosting services and internet service providers. DIGI has also engaged extensively with the Office of eSafety Commissioner on the development of their draft standards for Designated Internet Services and Relevant Electronic Services, which were registered on 21 June 2024, and will come into effect on 21 December 2024. 2.7. The AVM Act, detailed above, covers CSAM depicting rape or torture, which has been the subject of 98% of notices served under the Act

3. Non-consensual sharing of intimate imagery	3.1. DIGI's relevant members have strict policies that do not allow the sharing of non-consensual intimate images, and work to rapidly remove these. 3.2. These policies form part of broader policies to remove content that promotes sexual violence, sexual assault or sexual exploitation. 3.3. As with other policy areas described above, these policies are enforced through a combination of human review, proactive machine learning technology and enforcement teams. 4. Some platforms have also introduced preventative measures that use image hashing technology to prevent the spread of known image-based abuse images, in order to prevent the reliance on user reporting. For example, StopNCII.org protects adults across the world by preventing their intimate images from being shared across participating platforms.	4.1. The OSA includes a removal scheme where people who are the victims of the sharing of non-consensual intimate images may complain to the Commissioner if online service providers have failed to act on reports to them. The Office can direct a request for removal to the social media service, and the service must remove the content within 24 hours. 4.2. A core expectation of the draft Basic Online Safety Expectations (BOSE) is that a provider of a service must take reasonable steps to minimise the extent to which non-consensual intimate images are available, and to make reports about the provider's related activities available to the Commissioner.
5. Minors' access to pornography and other age-inappropriate content	5.1. All members have strict content policies in relation to pornographic content. On social media and content platforms, there are policies in their community guidelines restricting nudity, pornography and sexually explicit content. On search engines, sexual and violent terms are removed from auto-complete and pornography is demoted (and in some instances, removed or blurred entirely) in search results unless the user is clearly searching for it. These policies are enforced through a combination of human moderation and machine learning that detects high numbers of flesh coloured pixels. 5.2. These policies are also reflected in members' advertising policies. For example, Google Search	5.5. DIGI is currently consulting with the Office of the eSafety Commissioner concerning the development of Phase 2 industry codes to regulate materials that are unsuitable for children and young people under 18 years old under the National Classification Scheme. 5.6. In 2024, the Government also announced an age assurance trial, encompassing both age verification and age estimation technologies, to explore their efficacy in protecting children from encountering pornography and other high-impact online content.

	does not allow hyperlinks that drive traffic to commercial pornography sites, nor does it allow pornography ads to be placed within its search engine, nor does it run Google ads against pornographic websites. On social media and content platforms, all members have strict controls on pornography, adult products and services, and nudity. 5.3. Relevant members set age restrictions on their user-generated content platforms and many other products to limit and discourage the use of services by underage users, ranging from under 13 to 18 as appropriate to the service. When a notice or express admission that a user is underage is received, it will be investigated and accounts will be suspended accordingly. Some services will also take steps to prevent users lying about their age to access an account after it has been denied, by placing a persistent cookie on the device to prevent the child from attempting to circumvent the age restriction or by using artificial intelligence to understand the true age of a user. 5.4. Relevant DIGI members have extensive programs in place to protect young people on their services. At the service provider level, they provide applications to enable family sharing and limitations on minors' devices, that include controlling their privacy settings, filtering, screen time limits and other features designed to safeguard minors' privacy and experiences online. At the search engine level, they filter ads containing or promoting nudity, sexually suggestive content, adult entertainment and other services from appearing within search results. At the app distribution level, restricted profiles can be	

	established where more mature content can be filtered out of the app store. At the browser level, parents can create restricted profiles for minors that allow parents to block and approve sites viewed, and where "safe search" is on by default in such accounts. At the platform level, there are similar "safe search" settings that hide sensitive content and remove blocked and muted accounts. There are also default privacy settings for minors, and additional safety measures for users in this category, including restrictions aimed at inappropriate interactions and CSAM material, as well as advertising restrictions.	
6. Advocacy of suicide and self-harm	6.1. All relevant DIGI members have policies prohibiting the advocacy of suicide and other self-harm. These policies extend beyond the rapid removal of such content, but aim to provide those at risk with links to services that may assist them. For example, searches relating to suicide on platforms link to Lifeline and other relevant support organisations. Flags for suicide and self injury are escalated and addressed with urgency. 6.2. Relevant larger platforms partner with mental health organisations in Australia to produce or promote a range of training and other support resources. 6.3. Such policies and partnerships also extend to material that glorify eating disorders such as anorexia nervosa, and bulimia.	6.4. Australia was the first country to criminalise pro-suicide websites in 2006 through the Criminal Code Amendment (Suicide Related Material Offences) Act 2005.

7. Advertising of illegal and potentially harmful goods and services	7.1. Relevant DIGI members have broad-ranging advertising policies that prohibit or restrict a long list of illegal and potentially harmful goods and services. These policies are adapted to jurisdictions including Australian law. These policies include, but are not limited to, topic areas such as online wagering, adult goods and services, alcohol and tobacco sales. 7.2. These policies include the prohibition of deceptive, misleading, or harmful business propositions, including restrictions on misleading, false, or unsubstantiated claims during the promotion of a product or service. 7.3. They also have varying restrictions on political advertising, and work with Federal, State and Territory electoral offices to prevent electoral interference, as well as more traditional electoral offences. 7.4. Furthermore, there are restrictions on discrimination in the targeting of advertising to prevent discriminating against legally protected categories of customers. 7.5. Members work hard to ensure that age-regulated advertising content, such as those for alcohol, are not served to minors. 7.6. Advertising requires pre-registration and is reviewed and approved before publishing, and non-compliant ads may be disproved or removed, and repeat offender accounts may be suspended.	7.7. Australian Consumer Law applies to digital platforms, and has prohibitions on false and misleading content, unfair contract terms and provisions relating to consumer guarantees, product safety. This law is administered by the ACCC and and the State and Territory consumer protection agencies. 7.8. In relation to online gambling, the ACMA administers the Broadcasting Services (Online Content Service Provider Rules) 2018 (the Rules). The Rules apply to online content service providers who provide gambling promotional content on online content services in conjunction with live coverage of a sporting event. 7.9. There are state and federal electoral laws that apply to digital content. In 2022, DIGI worked with representatives from the Electoral Council of Australia and New Zealand (ECANZ) to establish escalation processes with platforms for federal elections.

8. Scams, spam and deceptive conduct	8.1. As well as the restrictions on advertising content, relevant members also have restrictions on organic as well as paid content in relation to scams, spam, fraud and other deceptive conduct. This includes phishing, impersonation and misrepresentation. 8.2. All of the measures outlined above from 2.1 to 2.4 (policies, tools, enforcement teams and technology) apply to the approach to scams, spam and deceptive conduct. 8.3. Relevant DIGI members are parties to the National Anti Scam Centre and its working groups, spearheading cross-sectoral initiatives to combat scam activity.	8.4. Australian Consumer Law applies to digital platforms, and has prohibitions on false and misleading content 8.5. The Australian Competition and Consumer Commission (ACCC)'s Scamwatch program enables consumers to complain to the ACCC that take action where appropriate, including working with industry. Scamwatch provides information to consumers and small businesses about how to recognise, avoid and report scams. State and Territory consumer protection agencies also have reporting and educational functions.[58] 8.6. DIGI notes there are forthcoming mandatory industry codes to target scam activity.
9. Privacy intrusion, hacking & threats to cyber security	9.1. DIGI's members have made and continue to make extensive investments in the privacy and safety of their users. At a high level, that work extends far beyond the provision of privacy policies, and includes notifications and privacy communication. Many provide privacy tools to provide people with transparency, choices and control about how their data is used. They have dedicated teams focused on privacy and cross-functional review processes for new products to ensure "privacy-by-design" before they are released. 9.2. Where applicable, they apply the strictest default privacy settings for minors; for example, ensuring that location-sharing is always off by default. 9.3. DIGI members all allow their users to destroy, de-identify, access and correct their personal	9.5. The Privacy Act and the Australian Privacy Principles apply to digital platforms, and DIGI welcomes the current review of these being led by the Attorney General's Department. We see this review as an important opportunity to standardise privacy protections in a digitising economy, and to ensure consumers have a baseline expectation of control and choice when it comes to their privacy. 9.6. DIGI notes major amendments to the *Privacy Act 1988* are expected to be introduced in 2024, including in relation to online privacy for minors.

[58] NSW Fair Trading, "Scams and cybercrime", accessed at https://www.fairtrading.nsw.gov.au/buying-products-and-services/scams

	information in accordance with the Australian Privacy Act 1988. 9.4. Their work in privacy is complemented by extensive investments in the cyber security of their users, which often includes the use of end-to-end encryption.	
10. Cyberbullying material targeted at an Australian adult	10.1. All of the measures outlined above (policies, tools, enforcement teams and technology) apply to the approach to cyberbullying material targeted at an Australian adult. 10.2. Digital platforms often have granular considerations when assessing the cyberbullying of adults, such as whether the content concerns public opinions or actions that impact others, and the extent to which the content relates to a person in authority or a public figure. The questions a provider may ask will necessarily differ based on the service, and provide important checks and balances for platforms to appropriately consider the freedom of expression, and political communication, implications of a takedown decision.	10.3. The OSA includes an adult cyber-bullying scheme where Australian adults who are the victims of seriously harmful online abuse can complain to the Office, if the online service providers have failed to act on reports to them. The Office can direct a request for removal to the social media service, and the service must remove the content within 24 hours. 10.4. The BOSE and Section 474.17(1) of the Criminal Code 1995 (Cth) detailed above also apply to adult cyberbullying.
11. Defamation	11.1. Relevant DIGI members have policies that restrict the usage of their services for the defamation of others. 11.2. They have complaints handling processes in place to action defamation requests received by Australian users, which are actioned in accordance with Australian law. These policies seek to balance allowing individuals to protect their reputations without placing unreasonable limits on the discussion of matters of public interest and	11.3. Defamation laws differ by state and territory in Australia, however the Model Defamation Provisions have played an important role in harmonising state-based defamation laws. The Council of Attorneys-General Defamation Working Party on the Review of Model Defamation Provisions (MDPs) advanced a process to ensure these provisions are fit for a digital age. 11.4. The reformed provisions require that digital intermediaries have in place a complaints mechanism that is reasonably accessible to the public, and where a complaints notice is

	importance. Given that defamation is a civil matter and can depend on whether the originator of a comment has a lawful defense for posting the comment, it can be challenging for platforms to make assessments in the absence of judicial or independent determinations.	lodged, to take certain steps within 14 days in order to preserve the availability of an innocent dissemination defence. The commencement date in NSW and ACT of the amendments to the Model Defamation Provisions are July 1, 2024.
12. Hate speech	12.1. All relevant DIGI members have strict policies to prohibit and address hate speech or conduct, which is generally defined as speech that maligns people or a group of people based on their protected characteristics, e.g. race, gender, sexuality. 12.2. These policies have and continue to evolve to capture emerging patterns and themes in hate speech or hateful conduct. Additionally, relevant members consult with a wide range of organisations and individuals who guide them in their policy decisions. 12.3. All of the measures outlined (e.g. policies, tools, enforcement teams and technology) apply to the approach to hate speech. 12.4. Industry policies and enforcement are complemented with a range of initiatives, partnerships and social programs aimed at preventing and addressing hate speech.	12.5. DIGI members take the aforementioned actions on hate speech under their own policies, despite no explicit and comprehensive legal protections for Australians under Australian law for hate speech. 12.6. There is a forthcoming hate speech bill to be introduced this year to provide further legal protections.

13. Misinformation and disinformation	13.1. Relevant DIGI members have policies and processes to remove or otherwise address the spread and scale of harmful misinformation and disinformation online, including user reporting mechanisms. As with other policy areas described above, these policies are enforced through a combination of human review, proactive machine learning technology and enforcement teams. 13.2. To provide a public, consistent and transparent framework for addressing the harm of mis- and disinformation to Australians, in February 2021, DIGI launched the *Australian Code of Practice on Disinformation and Misinformation* (ACPDM). 13.3. Eight major technology companies have adopted the code to date, and signatories have agreed to safeguards to protect Australians from harmful misinformation online. That includes the mandatory commitment (#1) of: 13.3.1.1. Publishing and implementing policies on their approach. 13.3.1.2. Providing a way for their users to report content that may violate those policies. 13.3.1.3. Implementing a range of scalable measures that reduce its spread and visibility online. 13.4. Another mandatory commitment (#7) is releasing annual transparency reports about those safeguards in order to improve public understanding of these challenges over time. The first set of reports were released in May 2021, and are available for anyone to read at digi.org.au. 13.5. The code contains opt-in commitments that have	13.8. DIGI developed the ACPDM in response to Australian Government policy announced in December 2019: "*The Government will ask the major digital platforms to develop a voluntary code (or codes) of conduct for disinformation and news quality. The Australian Communications and Media Authority (ACMA) will have oversight of the codes and report to Government on the adequacy of platforms' measures and the broader impacts of disinformation. The codes will address concerns regarding disinformation and credibility signalling for news content and outline what the platforms will do to tackle disinformation on their services and support the ability of Australians to discern the quality of news and information. The codes will be informed by learnings of international examples, such as the European Union Code of Practice on Disinformation. The Government will assess the success of the codes and consider the need* 13.9. DIGI supports the ACMA having greater powers as a regulator in relation to misinformation; a comprehensive approach that includes restrictions and oversight on misinformation across different industries that the ACMA oversees will ensure an ecosystem approach to misinformation. 13.10. DIGI notes that a revised *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* will be introduced to parliament this year.

	been widely adopted that entail (#2) Addressing disinformation in paid content. (#3) Addressing fake bots and accounts.(#4) Transparency about source of content in news and factual information (e.g. promotion of media literacy, partnerships with fact-checkers) and (#5) political advertising and (#6) partnering with universities/researchers to improve understanding. 13.6. In October 2021, DIGI announced the strengthening of the code with the appointment of an independent Complaints Sub-Committee comprised of Dr Anne Kruger, Victoria Rubensohn AM and Christopher Zinn to resolve complaints about possible breaches by signatories of their code commitments. DIGI launched a portal on its website for the public to raise such complaints. 13.7. In addition, DIGI appointed an independent expert Hal Crawford to fact check and attest signatories' annual transparency reports going forward under the code, in order to incentivise best practice and compliance.[59]	

Appendix 2 - Active government processes related to digital harms currently underway

1. The statutory review (review) of the Online Safety Act 2021 (OSA)
2. Recent Online Safety (Basic Online Safety Expectations) Amendment Determination;
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;

[59] DIGI Media Release (11/10/21), "Australian disinformation code of practice strengthened with independent oversight and public complaints facility", accessed at https://digi.org.au/in-the-media/australian-disinformation-code-of-practice-strengthened-with-independent-oversight-and-public-complaints-facility/

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. Misinformation and Disinformation Bill 2023 and associated processes;

10. Voluntary code for online dating services;

11. Voluntary AI Safety Standard;

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. 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.