



Committee Chair  
Select Committee on Social Media and Online Safety  
By email: [smos.reps@aph.gov.au](mailto:smos.reps@aph.gov.au)

Wednesday January 12, 2022

Dear Mrs Lucy Wicks MP,

The Digital Industry Group Inc. (DIGI) thanks you for the opportunity to provide a submission to the House Select Committee (the Committee) Inquiry on Social Media and Online Safety.

By way of background, DIGI is a non-profit industry association that advocates for the interests of the digital industry in Australia. DIGI's founding members are Apple, eBay, Google, Linktree, Meta, Twitter, Snap and Yahoo, and its associate members are Change.org, Gofundme, ProductReview.com.au and Redbubble. DIGI's vision is a thriving Australian digitally-enabled economy that fosters innovation, a growing selection of digital products and services, and where online safety and privacy are protected.

DIGI shares the Government's strong commitment to combatting harmful content online. This is evidenced by the fact that DIGI developed *The Australian Code of Practice on Disinformation and Misinformation* (ACPDM) to realise Australian Government policy in this area. Signed by eight major technology companies and open to any others, every code signatory commits to safeguards to protect against online mis- and disinformation, including publishing and implementing policies on their approach, and providing a way for their users to report content that may violate those policies. Signatories must release annual transparency reports about all of those efforts, the first set of which were released in May 2021, providing new insights on the scale of the online misinformation in Australia and its management.

DIGI is also currently working with a wide range of companies, well beyond our membership, to develop industry-wide mandatory codes under the Online Safety Act 2021 (OSA Codes). We are working with the Office of the eSafety Commissioner (the Office) to have the first set of these registered around July 2022, and they will be released for public consultation prior to then. DIGI is the drafting of the chapters of the codes relating to social media services, search engines, and app distribution services. Once in effect, these codes will standardise industry-wide protections for Australians in relation to Class 1 and Class 2 content under the classification code, which includes child sexual exploitation material, pro-terror content and pornography.

DIGI has also worked on the prevention of online harms through our annual *DIGI Engage* youth summits. For four years, we have 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. The *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.

DIGI's work in these areas and others is a reflection of our members' deep commitment, longstanding and continued investment in online safety. Section 3 of this submission features a non-exhaustive, high-level overview of our relevant members' work in response to online harms, which includes policies, moderation, technology, partnerships and research engagement. DIGI's members believe in "safety by design", and many of them have worked with the Office to guide and endorse their assessment tools in this area. Those that operate social media services also release transparency reports that provide data and details of their approach to their enforcement of policies and laws.



DIGI welcomes this inquiry as a way to have a deeper conversation about online safety, surface challenges and identify action areas. From DIGI's vantage point, we see the extensive work our members undertake in addressing online harms, as well as the work that the industry does together. Online safety can never be a "set and forget" exercise; it is right that platforms' online safety measures and their impact are analysed and questioned to aid continual improvement.

DIGI and its members believe that the digital industry has a responsibility to address online harm, and that the Australian Government has an important role to play in standardising protections, encouraging accountability and providing safety nets for consumers. DIGI sees itself as a key Government partner in this endeavour, through the work outlined above, and our ongoing engagement with proposed regulation where we advocate for approaches that are effective in their goals and can practically be implemented by industry.

One of those areas of regulation is the Online Safety Act (OSA) which comes into force on January 23, 2022. This is broad-ranging reform that includes:

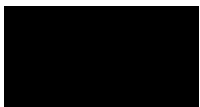
- New takedown schemes that require the removal of certain harmful and illegal content within 24 hours, including a new adult cyberbullying scheme;
- Basic Online Safety Expectations (BOSE), the draft of which establishes a set of basic safety standards and reporting requirements that go beyond the takedown schemes of the Act for all social media services, messaging services and websites.
- A new Restricted Access System (RAS) which requires that social media services, messaging services and websites limit access to certain age-inappropriate material.
- The aforementioned OSA codes, to be registered by the Office of the eSafety Commissioner later in 2022, that relate to "Class 1" and "Class 2" materials under Australia's classification code.

DIGI looks forward to engaging with this inquiry and advancing our shared goals around online safety. To that end, this submission provides:

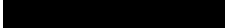
1. A high-level overview of the range of online harms, and brief summary of industry and regulatory approaches to each, which underpins the discussion and recommendation in other sections, provided in **Section 3**;
2. An exploration of particular issues relating to online harms that DIGI expects may be of interest to the Committee, in **Section 2**;
3. Advancement of specific recommendations that we encourage the committee to consider in its efforts to advance online safety for Australians in 2022, **which we open with in Section 1**.

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

Best regards,



Sunita Bose  
Managing Director, DIGI





## Table of contents

<b>Section 1: Recommendations</b>	<b>4</b>
#1 Streamlining online safety legislation into one consolidated act	4
#2 End-user notices for online safety breaches	5
#3 Comprehensive legal standards for hate speech and hate groups	5
#4 Strengthened safety nets for the disinformation and misinformation code	6
#5 Economy-wide protections for minors' data in Privacy Act Review	6
#6 Streamlining age verification regulatory processes to better explore Australians' attitudes to it	7
#7 A targeted approach to protecting minors from age inappropriate content	8
#8 Evidence-informed approach to youth mental health and social media	8
<b>Section 2: Areas of potential interest to the Committee</b>	<b>9</b>
Section 2a: Protecting minors online	9
Rapid removal of harmful content	9
Protections for children's privacy	9
Protections from pornography and age-inappropriate content	9
Age assurance	11
Youth mental health and social media	13
Section 2b: Other issues not specific to minors	14
Anonymity & pseudonymity online	14
Algorithms and online harms	17
Public figures and private citizens	19
<b>Section 3: Online harms on social media and online platforms</b>	<b>20</b>
Overview of online harms, platform & regulatory responses	20
Cyberbullying material directed at an Australian child	22
Cyberbullying material targeted at an Australian adult	23
Hate speech	23
Defamation	24
Misinformation and disinformation	25
Pro-terror material and the incitement of violence	26
Child sexual abuse material (CSAM)	28
Non-consensual sharing of intimate imagery	29
Minors' access to pornography and other age-inappropriate content	29
Advocacy of suicide and self-harm	31
Advertising of illegal and potentially harmful goods and services	31
Scams, spam and deceptive conduct	32
Privacy intrusion, hacking & threats to cyber security	33
Shared challenges in addressing online harms	34
Conclusion	35



## Section 1: Recommendations

This section provides a summary of DIGI's key recommendations made in this submission that we encourage the committee to consider in its efforts to advance online safety in Australia in 2022. The recommendations are underpinned by gaps identified in the overview provided in Section 3.

### 1. #1 Streamlining online safety legislation into one consolidated act

- 1.1. DIGI was very supportive of efforts to streamline legislation pertaining to online safety under one consolidated Online Safety Act (OSA). In Table 1 in Section 3 of this submission, we provide a high-level overview of the range of online harms, and a brief summary of industry and regulatory approaches to each. Table 1 serves to illustrate the wide range of laws and initiatives underway pertaining to online regulation and demonstrates the complexity of the regulatory environment.
- 1.2. While the OSA does, to a large extent, achieve its aim of consolidating the law relating to online safety, further consolidation may be helpful. For example, how service providers are expected to handle abhorrent violent content is covered under several different schemes in the OSA, including its takedown schemes, the OSA codes, and the BOSE. However, this same subject matter is *also* covered under the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (AVM Act).
- 1.3. Despite the effort made to streamline online safety legislation into the OSA, we are seeing inconsistencies emerge across the various online regulatory instruments that are being progressed under it. As one example of an inconsistency, the OSA's takedown schemes and the BOSE suggest that service providers should be required to remove all types of Class 1 material. However, the Commissioner's position as stated in their position paper on the OSA Codes<sup>1</sup> is that an identified subclass of Class 1, termed "Class 1b (fetish practices)" can be treated as Class 2 materials, and therefore do not need to be removed. It is unclear whether this interpretation extends to other aspects of the OSA, which creates confusion for industry participants working in good faith to comply with the legislation.
- 1.4. Streamlining online safety legislation into a consistent and consolidated Act will aid clarity and compliance – particularly for start-ups, smaller challenger companies, and those without a large local staff presence – who may be struggling to make sense of the complex regulatory environment in Australia. Developing a comprehensive regulatory response to online safety through one overarching piece of legislation is the approach being taken in the European Union through their Digital Services Act and in the UK through the Online Safety Bill. This is particularly important when we consider the broad scope of the OSA and other related online safety legislation, which often applies to all websites and/or services enabling interaction in Australia (i.e. not just large social media services).

---

<sup>1</sup> Office of the eSafety Commissioner, *Development of industry codes under the Online Safety Act: Position Paper*, accessed at <https://www.esafety.gov.au/about-us/consultation-cooperation/industry-codes-position-paper>.



## 2. #2 End-user notices for online safety breaches

- 2.1. DIGI welcomes the inclusion of end-user removal notices in the OSA whereby end-users, who post certain restricted material under that act, may be given a notice from the eSafety Commissioner requiring them to remove the material. To the extent that such notices serve to discourage the posting of further abusive content on multiple platforms – or if their scope of the OSA's notices might be expanded to this end – such an approach may serve to deter the end user from posting material on different providers' services. This may also prevent the need for victims of abuse to have to report multiple pieces of content from a single perpetrator to multiple platforms.
- 2.2. We encourage the Government's consideration of more behavioural and perpetrator level policy approaches to complement platform-level takedown schemes. In relation to end users, the Australian Government made an election commitment May 5, 2019 to increase maximum penalties for end-users who use a carriage service to menace, harass or cause offence to five years of imprisonment<sup>2</sup>.
- 2.3. It is worth noting that the current Enhancing Online Safety Act (EOSA) scheme enables the Office 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; yet, to date, we understand that no such end-user notices relating to cyberbullying have been issued in the six years of its operation. With the expansion of the takedown schemes to include adult cyberbullying under the OSA, we encourage further strategic emphasis on the issuance of end-user notices in the Office's regulatory approach.

## 3. #3 Comprehensive legal standards for hate speech and hate groups

- 3.1. While DIGI's relevant members all have long-standing policies prohibiting hate speech, DIGI is concerned that there is no legal standard nor recourse in this area to incentivise strong action industry-wide. Australia continues to adopt a narrow approach to hate speech under anti-discrimination laws and the OSA that are aimed at protecting individuals, rather than groups based on their protected characteristics. DIGI believes that hate speech should be a broader category of harm that is not limited to religious, gender or race-based discrimination, but includes hate speech related to characteristics including (but not limited to) sexuality, gender identity, disability, and national origin.
- 3.2. DIGI has and continues to encourage the Australian Government to develop a clearer legislative framework that encompasses this broader approach to hate speech to assist enforcement agencies and prosecutors<sup>3</sup>. This will also serve to help relevant stakeholders, including all digital

---

<sup>2</sup> See media release: Prime Minister The Hon Scott Morrison MP, Attorney-General The Hon Christian Porter, Senator The Hon Mitch Fifield Minister For Communications And The Arts, Joint Media Release (05/05/2019), "Keeping Australians Safe Online", accessed at <https://www.liberal.org.au/latest-news/2019/05/05/keeping-australians-safe-online>. See also transcript: Prime Minister The Hon. Scott Morrison MP (05/05/2019), Transcript Remarks, Campaign Rally Central Coast, accessed via CCH alerts, see quote: "But the other thing we're going to do for all Australians, is we're going to increase the penalties for those who have been found to be bullying people online, causing those injuries. You won't go to jail for three years you'll go to jail for five years."

<sup>3</sup> DIGI, Submission to Department of Communications on the Online Safety Charter (14/04/09), accessed at <https://digi.org.au/advocacy/#:~:text=Online%20Safety%20Charter%20%7C%20Submission%20to%20Department%20of%20C,ommunications>



platforms, to better report, review and remove content that meets a defined Australian legal threshold.

- 3.3. Additionally, the Australian Government may also provide further legal clarity by reviewing the protocol for listing terrorist organisations<sup>4</sup> in response to the growing threat from the far right and consider whether new organisations should be added to further efforts to address hate. This might be similar to the FBI list of Foreign Terrorist Organisations and the UK's list of proscribed terrorist groups<sup>5</sup>. Clear and comprehensive guidance will assist industry players seeking to promptly remove content from terrorist groups.

#### 4. #4 Strengthened safety nets for the disinformation and misinformation code

- 4.1. DIGI launched The Australian Code of Practice on Disinformation and Misinformation (ACPDM) in February 2021 to realise Australian Government policy in this area. In October 2021, DIGI announced the strengthening of the ACPDM through 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 also launched a portal on its website for the public to raise such complaints<sup>6</sup>.
- 4.2. DIGI is yet to see the ACMA's report on the effectiveness of the ACPDM that it provided to the Government on June 30, 2021. Reviewing this report will assist DIGI in its efforts to continue to strengthen the code in line with expectations. DIGI intends to conduct a review of the ACPDM in 2022.
- 4.3. In the absence of seeing this report, and in advance of the review, over the course of September-November 2021, DIGI has made a recommendation to the Australian Government (via the ACMA, Minister's Office and the Department of Communications) for how the ACPDM can be strengthened, for which we are awaiting an outcome. DIGI has presented an identified gap in the governance arrangements outlined should disagreements arise between the complainants of the ACPDM and the Complaints Sub-Committee and has proposed that the ACMA provide this appeals role. We believe an appeals process operated by the ACMA will provide an important safety net for consumers in relation to the ACPDM.

#### 5. #5 Economy-wide protections for minors' data in Privacy Act Review

- 5.1. DIGI fully supports the intention of the Online Privacy Bill (OPB) to protect the privacy of minors online and to safeguard them from harm. However, we do not believe that the scope of services covered under the Bill is wide enough to ensure a standardisation of protection for minors across all digital services they use; for example, it does not include education technology, health

---

<sup>4</sup> Australian Government, "Listed terrorist organisations", accessed at

<https://www.nationalsecurity.gov.au/what-australia-is-doing/terrorist-organisations/listed-terrorist-organisations>

<sup>5</sup> UK Government (12/07/2013, last updated 26/11/2021) "Proscribed terrorist groups or organisations", accessed at <https://www.gov.uk/government/publications/proscribed-terror-groups-or-organisations-2>

<sup>6</sup> 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/>



services, or the banking sector. We therefore encourage specific privacy protections for minors be expanded upon within the Privacy Act, as this Act covers all organisations with an annual turnover of more than \$3 million and Australian Government agencies<sup>7</sup>. The review of that Act provides an important opportunity to standardise and strengthen a single set of routine privacy protections that minors, or their guardians, can expect online.

- 5.2. DIGI recommends that the Privacy Act Review Discussion Paper proceeds with its recommendation outlined below to address this issue:

*APP entities that engage in the following restricted practices must take reasonable steps to identify privacy risks and implement measures to mitigate those risks:*

- *Direct marketing, including online targeted advertising on a large scale\**
- *The collection, use or disclosure of sensitive information on a large scale*
- *The collection, use or disclosure of children's personal information on a large scale*
- *The collection, use or disclosure of location data on a large scale*
- *The collection, use or disclosure of biometric or genetic data, including the use of facial recognition software*
- *The sale of personal information on a large scale*
- *The collection, use or disclosure of personal information for the purposes of influencing individuals' behaviour or decisions on a large scale*
- *The collection use or disclosure of personal information for the purposes of automated decision making with legal or significant effects, or*
- *Any collection, use or disclosure that is likely to result in a high privacy risk or risk of harm to an individual.*

*\*'Large scale' test sourced from GDPR Article 35. Commissioner-issued guidance could provide further clarification on what is likely to constitute a 'large scale' for each type of personal information handling.<sup>8</sup>*

## 6. #6 Streamlining age verification regulatory processes to better explore Australians' attitudes to it

- 6.1. DIGI notes that the age verification requirements were a new addition to the OPB when it was released on October 25, 2021 and had not previously been foreshadowed when the Bill was first announced on March 24, 2019 nor in the ensuing 2.5 year period. Given the unprecedented implications of age verification of Australians on a wide range of digital services, in light of the extremely broad definition of "social media services", wider consultation must take place in relation to this specific proposal.
- 6.2. DIGI is concerned that the OPB, expected to be introduced to parliament early in 2022, has not drawn Australians' attention to the fundamental changes that they are proposing to how Australians use the Internet. Do all adults and minors want to routinely provide their personal information, age data, and potentially identity verification documents when they are perusing

<sup>7</sup>Office of the Australian Information Commissioner (OAIC), "Who has rights under the Privacy Act?", accessed at <https://www.oaic.gov.au/privacy/the-privacy-act/rights-and-responsibilities#WhoHasResponsibilitiesUnderPrivacyAct>

<sup>8</sup> Attorney General's Department (25/10/21), *Privacy Act Review Discussion Paper*, accessed at: [https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user\\_uploads/privacy-act-review-discussion-paper.pdf](https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user_uploads/privacy-act-review-discussion-paper.pdf), p.97





websites that include user interaction? We believe that further consultation needs to occur on the propensity of Australians to undertake age verification and additional personal information provision as a pre-requirement to use digital services, before these requirements are potentially legislated. DIGI suggests that the Office of the eSafety Commissioner's existing Age Verification Roadmap provides an appropriate framework for this consultation to occur, which will also avoid duplication across multiple regulations and processes.

## 7. #7 A targeted approach to protecting minors from age inappropriate content

- 7.1. Instead of the parental consent and verification requirement being proposed by the Australian Government under the OPB, DIGI suggests a focus on other Government and industry measures in train that are aimed at minimising the exposure of age-inappropriate content to minors. In addition to the Age Verification Roadmap, the OSA Codes (as detailed in Section 2 of this submission) will be registered by the Office of the eSafety Commissioner in 2022, and are expected to cover in scope the tools available to parents to manage and oversee their children's experiences online. The first set of codes will be released by DIGI, Communications Alliance and other industry associations for public consultation in the first half of 2022.

## 8. #8 Evidence-informed approach to youth mental health and social media

- 8.1. Government recommendations made in relation to youth mental health and social media should, in DIGI's view, be evidence-informed and situated within a whole-of-Government approach, such as part of the response to the Productivity Commission's Inquiry into mental health in Australia.
- 8.2. While it is possible that there may be such research underway (noting that industry participants do routinely undertake or support research efforts, detailed in their own submissions), it is DIGI's understanding that research to date has not established a direct causal link between social media use and youth mental health issues in Australia, and that the existing research indicates a level of complexity in the interaction with other factors (e.g. social contexts and familial conditions in which children and young people live<sup>9</sup>). DIGI would be very supportive of further Australian research that aims to understand the link between the two, 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; In line with suggestions from a UNICEF research paper, researchers need to consider children's life contexts and socio-demographics to ensure that variables that have known effects on child well-being outcomes are not excluded. 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 industry would be open to collaborating on this effort.

---

<sup>9</sup> 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>





## Section 2: Areas of potential interest to the Committee

This section explores particular issues relating to online harms that DIGI expects may be of interest to the Committee.

### Section 2a: Protecting minors online

#### 9. Rapid removal of harmful content

- 9.1. DIGI believes there are several industry approaches and legal safety nets that are crucially important to protecting minors online, which we have mapped in detail in Section 3 of this submission.
- 9.2. In brief, this includes the rapid removal of relevant harmful content, enabled by strict policies and enforcement architecture to prohibit and rapidly remove the cyberbullying of Australian children and minors.
- 9.3. DIGI members have zero tolerance for child sexual exploitation materials (CSAM), rapidly removing it and working closely with Australian law enforcement to report and address it. Their extensive work in this area is summarised in Section 3.
- 9.4. There are takedown schemes under the OSA relating to cyberbullying, CSAM and a wide range of harmful content (detailed in Section 3), 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 – which violates major consumer platforms including DIGI members' policies – are not rapidly removed as intended. **Per Recommendation #2, where victims of cyberbullying in particular are experiencing bullying on multiple platforms, in order to prevent the need for victims of abuse to have to report multiple pieces of content from a single perpetrator to multiple platforms, we encourage further strategic emphasis on the issuance of end-user notices in the Office's regulatory approach.**

#### 10. Protections for children's privacy

- 10.1. In addition, and as explored above, there needs to be widespread protections for children's privacy and their data. As detailed in Section 1, DIGI sees a key opportunity to standardise these protections in the current Privacy Act Review in order to provide minors, or their guardians, confidence that a baseline standard of privacy exists no matter which online service they are using. **Per Recommendation #5, DIGI encourages specific privacy protections for minors be expanded upon within the Privacy Act.**

#### 11. Protections from pornography and age-inappropriate content

- 11.1. Design features and tools are important to protect minors from age-inappropriate content, such as pornography. One way to ensure parents can make decisions based on the evolving capabilities of minors 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;



DIGI members have extensive experience in developing and implementing such controls. At the service provider level, Apple<sup>10</sup> and Google<sup>11</sup> provide applications to enable family sharing and limitations on minors' phones and tablets, that include controlling their privacy settings, filtering access to content, screen time limits and other features designed to safeguard minors' privacy and experiences online. At the search engine level, Google's Safe Search filter<sup>12</sup> prevents search results 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. 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. 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; for example, Instagram defaults users between the ages of 13 and 17 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<sup>13</sup>. On Snapchat, default settings for *all users* prevent receiving a message from someone who is not your friend and location sharing is off by default, and there is no option for users to share location outside of their friend group. Platforms 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<sup>14</sup>.

- 11.2. The forthcoming OSA codes, to be registered by the Office of the eSafety Commissioner in 2022, will include measures designed to protect minors from certain age inappropriate materials, consistent with the National Classification Scheme. DIGI and the Communications Alliance, supported by a steering group of other industry associations, are developing these codes to regulate all online services and websites available in Australia, which will be registered by the Office of the eSafety Commissioner in 2022.
- 11.3. At a high level, the OSA codes will contain commitments from industry to implement measures that minimise the risk of harm to Australian end-users that may be caused by Class 1 materials online. The OSA Codes will also contain commitments from industry to minimise the risk of harm to Australian minors due to the accessibility of Class 2 materials online. Class 1 materials include child sexual exploitation material, pro-terror content, content that depicts, promotes, incites or instructs in matters of crime, violence or drug misuse, and online pornography that depicts fetish practices or fantasies. Class 2 materials include other online pornography, X18+ and R18+ content, and material which includes high-impact sex, nudity, violence, drug use, language and

<sup>10</sup> Apple Support, *Use parental controls on your child's iPhone, iPad, and iPod touch*, accessed at: <https://support.apple.com/en-us/HT201304>

<sup>11</sup> Google, *Google Family*, accessed at: <https://families.google.com/familylink/>

<sup>12</sup> 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.

<sup>13</sup> See the following links for more information: Youtube, *You Tube - 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>

<sup>14</sup> 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/>



themes; 'Themes' includes social Issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism..

- 11.4. The OSA codes will apply to eight sections of the industry, namely: providers of social media services (defined around online social interaction between 2 or more end-users), providers of relevant electronic services (includes any services with messaging, and gaming), providers of designated internet services (includes all websites), providers of internet search engine services, providers of app distribution services, providers of hosting services, providers of internet carriage services, and persons who manufacture, supply, maintain or install certain equipment (includes retailers). This is broader in scope than the OPB's requirements pertaining to age and guardian verification, which are limited to social media services. The OSA codes therefore provide a "whole of industry" approach to the issue of minors' access to age-inappropriate content.
- 11.5. DIGI and other industry associations involved expect to release the first set of codes for public consultation in the first half of 2022, with a view to having the first set registered by July 2022; the remaining codes will be released for public consultation and registered by the end of the year. **Per Recommendation #7, we believe the OSA codes should provide a targeted and systemic approach to protecting minors from age-inappropriate content.**

## 12. Age assurance

- 12.1. DIGI is supportive of the usage of *age assurance*, as opposed to *age verification*, in enforcing age restrictions on services and to otherwise prevent the exposure of age-inappropriate content to minors.
- 12.2. We are concerned that the proposed OPB has adopted the narrower language of "age verification" rather than the broader spectrum of "age assurance" solutions. *Age verification* is the most privacy-intrusive form of *age assurance*. If the OPB's requirement to "take all reasonable steps to verify the age of individuals" equates to identity verification, through the provision of drivers' licences, passports or other Government issued identification documents, the privacy intrusion of the Bill will be immense.
- 12.3. Australia's proposals for widespread age verification are inconsistent with approaches to these challenges in the EU and the UK. The EU's Audio Visual Media Services Directive (AVMSD) contemplates age verification as a possible measure "for users of video-sharing platforms with respect to content which may impair the physical, mental or moral development of minors"<sup>15</sup>. It does not contemplate the universal application to broadly scoped "social media services" that enable user interaction.
- 12.4. Nor does the UK's Age Appropriate Design Code (AADC), which provides a useful model when considering policy approaches to this complex issue. This code came into force on September 2, 2020 with a 12 month transition period where organisations must confirm by September 2, 2021. The code sets out 15 standards, and it is noteworthy that the standards do not include age verification, but rather elevate the standard of data minimisation. On the topic of age assurance, the AADC guidance material states:

*We recognise there is a tension between age assurance and compliance with GDPR, as the implementation of age assurance could increase the risk of intrusive data collection. We do*

<sup>15</sup> European Union (2018), *Audio Visual Media Services Directive*, accessed at <https://eur-lex.europa.eu/eli/dir/2018/1808/oj>



*not require organisations to create these counter risks. However, age assurance and GDPR are compatible if privacy by design solutions are used<sup>16</sup>.*

- 12.5. In April 2021, the United Nations Children’s Fund (UNICEF) released a discussion paper, titled “Digital Age Assurance, Age Verification Tools, and Children’s Rights Online across the Globe: A Discussion Paper”<sup>17</sup>, which examines age assurance tools and provides a useful account of the available methods. This discussion paper identifies the main potential data sources and methods for age assurance tools, which are: State or government-provided, user-provided data such as official documents, automatically generated data, biometric data, blockchain and self-sovereign identities, behavioural and self declaration.
- 12.6. In its assessment of each of these age assurance methods, there are still privacy concerns about the additional collection of personal information associated with all. The UNICEF report examines each age assurance approach against the United Nations Convention of the Rights of the Child (CRC). Article 16 of the CRC focuses on the right to privacy states:

*Most age assurance tools with a high degree of accuracy rely on official data that can easily identify a child. It is important that children’s right to privacy is respected as they continue to engage in online spaces, and that they are only identified where strictly necessary to prevent serious harm, and with their consent or the consent of their parents or caregivers<sup>18</sup>.*

- 12.7. The paper raises the need for age assurance processes to respect the data minimisation principle, and this is one of the fundamental challenges related to this proposal where we need a balanced approach. DIGI notes that self-declaration is the most common age assurance method used by the digital industry, which optimises for data minimisation and usability, at the risk of accuracy.
- 12.8. The privacy intrusion of the OPB is multiplied by the requirement to obtain parental or guardian express consent before collecting information. Not only could this require additional collection of personal information from additional end users, it may require the collection of secondary documents to verify parental status or guardianship. For example, there are many parents or guardians who do not have the same last name as their children. This may be because their children have the last name of their spouse, due to adoption, or for legal guardians who are not biological parents. Is the Government expecting parents to provide a birth certificate, Medicare card or other identification in order to demonstrate their guardianship of a particular minor? This would increase the amassing of personal identification documents at the platform level.
- 12.9. Neither the OPB, nor its Explanatory Paper, contemplate scenarios where a young adult may require access to digital services outside of the purview of their legal guardian, such as to access assistance or health information. This is particularly relevant in situations where the minor’s relationship with their guardian may be constrained.

---

<sup>16</sup>UK Information Commissioner’s Office (2020), *Age appropriate design: a code of practice*, accessed at <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf>, p. 35

<sup>17</sup>UNICEF, (2021), *Digital Age Assurance, Age Verification Tools, and Children’s Rights Online across the Globe: A Discussion Paper*, accessed at <https://c-fam.org/wp-content/uploads/Digital-Age-Assurance-Tools-and-Childrens-Rights-Online-across-the-Globe.pdf>

<sup>18</sup> UNICEF, p.11



- 12.10. This is one of several concerns identified in the UNICEF Discussion Paper in relation to the CRC's Article 2 concerning non-discrimination, where it states:
- It is important that age assurance processes do not inadvertently discriminate against children who do not have access to official documents, children with developmental delays, children whose ethnicity is not recognized by algorithms used to assess age, or children who do not have parents or caregivers who are able to engage with verification processes that require parental input.*
- 12.11. The UNICEF discussion paper also discusses Article 5 of the CRC, which focuses on parental guidance and a child's evolving capabilities where it states: "It may be difficult to reconcile age-based restrictions with the concept of the evolving capacities of the child." Consideration needs to be given to the varying impact of the Bill on young adults, not just children.
- 12.12. **Per Recommendation #6, DIGI recommends further consultation needs to occur on the propensity of Australians to undertake age verification and additional personal information provision as a pre-requirement to use digital services, before these requirements are potentially legislated.**
- 12.13. **Per Recommendation #7, instead of the parental consent and verification requirement being proposed by the Australian Government under the OPB, DIGI suggests a focus on other Government and industry measures in train that are aimed at minimising the exposure of age-inappropriate content to minors such as the OSA Codes and Age Verification Roadmap.**

### 13. Youth mental health and social media

- 13.1. DIGI recognises the Committee and community concerns in relation to youth mental health and social media, 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.
- 13.2. For its part, DIGI has engaged with research in this area through its position on the advisory board of the University of Western Sydney's Young and Resilient Centre, that works with children and young people globally researching the role of technology to inform policies, programs and interventions that can minimise the risks and maximise the benefits of the digital age. Research by the centre's co directors, titled *Social Media and the Wellbeing of Children and Young People: A Literature Review*<sup>19</sup> concludes:

*This review provides a high level snapshot of the evidence of children and young people's social media use and the effects on wellbeing. It shows the effects are broadly positive, but are mediated by the social contexts and familial conditions in which children and young people live. Children and young people bring to their social media use pre-existing social, cultural, political, emotional and psychological experiences and status. It is the ways in which they interact with social media to produce identity, community and culture that provide the clearest insight into the role of social media for wellbeing. Moreover, how policy-makers, carers, professionals and service providers respond to social media in policy, service delivery and practice all contribute to the broader debates and practices by which social media affects the wellbeing of children and young people. The benefits and*

<sup>19</sup> 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>



*risks of social media use map to broader patterns of communicative and literacy practices, as well as socio-economic and cultural disadvantage. Intervening in this cycle has the potential to generate a steep change in the wellbeing of the children and young people who stand most to benefit. Such efforts must not only be informed by research, but by the views and preferences of children and young people themselves.*

- 13.3. While it is our understanding that a direct causal link between social media use and youth mental health issues in Australia has not been established to date through available research, **per Recommendation #8, we would be very supportive of further Australian research that aims to understand the link between the two 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 Kardefelt-Winther argues in a UNICEF Office of Research paper, children's online experiences cannot be studied in isolation from their lives in general.

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

- 13.4. None of that is to imply that research is required as precursor to action; DIGI members conduct significant work related to mental health in a range of ways. They have "downstream" 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. They also invest "upstream" through partnerships to deliver youth resilience-building programs, in areas such as cyberbullying and digital literacy and anti-bullying.

## Section 2b: Other issues not specific to minors

### 14. Anonymity & pseudonymity online

- 14.1. DIGI has observed that many of the Australian Government's recent policy proposals are moving toward the removal of anonymity and pseudonymity online. Some of these proposals appear to be premised on a hypothesis that online harms are correlated with anonymity and pseudonymity online. DIGI urges that this hypothesis be rigorously explored, along with an exploration of the unintended consequences of discouraging anonymity and pseudonymity.
- 14.2. The OPB, the Social Media (Anti-trolling) Bill 2021 (ATB) and the BOSE all apply to "social media services", defined broadly to encompass interaction between "two or more end users". This definition is by no means limited to large, mainstream social media services as it encompasses a wide range of services, such as local and small business community forums, educational forums, business forums, health support forums, and any blogs with comments enabled. For example, the mental health organisation Beyond Blue operates a number of online community forums on





topics relating to anxiety and depression where Australians can share their experiences and connect<sup>20</sup>.

14.3. The OPB has requirements for all such services to:

- *Take all reasonable steps to verify the age of individuals who use the social media service; and*
- *Ensure that the collection, use or disclosure of a child's personal information is fair and reasonable in the circumstances, with the best interests of the child being the primary consideration when determining what is fair and reasonable; and*
- *Obtain parental or guardian express consent before collecting, using or disclosing the personal information of a child who is under the age of 16, and take all reasonable steps to verify the consent. In the event that a social media service becomes aware that an individual was under the age of 16 (for instance if they had new information to suggest an individual previously believed to be over the age of 16 was in fact not), the social media service must take all reasonable steps to obtain verifiable parental or guardian consent as soon as practicable.*

If the Bill's focus on age *verification* is retained, as opposed to age *assurance*, it could result in the widespread collection of identity verification documentation such as drivers' licences, and documents that prove guardianship, such as Medicare cards and birth certificates.

- 14.4. The ATB, which is currently out for consultation, adopts a similarly broad definition as the OPB and applies to websites available in Australia that enable interaction between two end-users. Note that DIGI will be providing a separate submission on the ATB and can provide this to the Committee once complete, if it is of interest.
- 14.5. The BOSE – which when finalised will come into effect with the OSA on January 23, 2022 – contains an expectation that service providers will take reasonable steps to prevent anonymous accounts from being used for unlawful or harmful materials or activities (Section 9 (1)). The steps service providers can take to meet this expectation include having processes that require verification of identity or ownership of accounts (Section 9(1) (b)).
- 14.6. Many end-users have valid and important reasons for anonymity. For example, as the Office of the eSafety Commissioner acknowledges, a valid reason for anonymity and identity shielding is to protect users from unwanted contact. The Office encourages children only to use their given name, a nickname or an avatar online instead of a full real name which makes it more difficult for sexual predators and scammers to interact with them<sup>21</sup>.
- 14.7. As acknowledged by the former UN Special Rapporteur for Human Rights David Kaye, the ability for users to remain anonymous online can also be an important means for keeping them safe and promoting human rights<sup>22</sup>. For example, anonymity enables activists to expose repression, corruption and hate. Anonymity allows stigmatised or marginalised communities to find safety and support when revealing their real-world identity could expose them to harm. As David Kaye recently told a forum in Australia on anonymity:

<sup>20</sup> Beyond Blue, *Online forums*, accessed at: <https://www.beyondblue.org.au/get-support/online-forums>

<sup>21</sup> Office of the eSafety Commissioner, "Anonymity and identity shielding", accessed at <https://www.esafety.gov.au/about-us/tech-trends-and-challenges/anonymity>

<sup>22</sup> Kaye, David (2015), *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Human Rights Council, accessed at <https://www.undocs.org/A/HRC/29/32>





*It's been essential to individual human development in repressive societies – the ability to seek information or receive information in a kind of cone of privacy, if we want to think of it like that, under the blanket of anonymity.*

*It has allowed people historically to explore their heritage, to explore their sexual orientation, their gender identity, and we could go on and on, and anybody could come up with examples where a failure of anonymity or publicity of one's persona might lead to real harm<sup>23</sup>.*

- 14.8. These regulatory proposals to collect additional information run counter to the universally accepted privacy practice of data minimisation. Data minimisation requires goods or service providers to not seek to collect data beyond what is reasonably needed to provide the good or service. Data minimisation that forms part of the existing APPs under the Privacy Act 1988 (cth)<sup>24</sup>, and is also a key principle of the Consumer Data Right<sup>25</sup>.
- 14.9. Additionally, we are concerned that a potential increase in data collection for all websites, and the sometimes sensitive nature of the data being collected, will create increased cyber security risks to users of a whole range of websites in Australia. A pertinent example is provided by the 2015 Ashley Madison data breach in the United States. In July 2015, user data was stolen from the company Ashley Madison, a commercial dating website associated with extramarital affairs, and threatened to be released if the company did not shut down. The following month, more than 60 gigabytes of company data was leaked, including user data such as real names, home addresses, search history and credit card transaction records<sup>26</sup>. It is a reasonable prediction that similar widespread attacks, intended to publicly shame users of certain websites through personally identifiable data, may occur if widespread age verification solutions are imposed.
- 14.10. In addition to the negative implications for Australians' safety, advocacy, privacy and cyber security, there is evidence that calls into question the correlation between anonymity and online harm. In August 2021, Twitter released an analysis of accounts that were removed or suspended for abuse on its platform in response to the Euro 2020 final, and found that 99% of the accounts suspended were not anonymous and were in fact identifiable<sup>27</sup>.
- 14.11. If Australia is to embark upon a regulatory approach to impose widespread identification of Australian Internet users, to our knowledge, it would be the only democratic nation to do so at this point in time. The only country, to our knowledge, that has legislation requiring the widespread

---

<sup>23</sup>Taylor, Josh (5/11/2021), "Twitter says any move by Australia to ban anonymous accounts would not reduce abuse", in *Guardian Australia*, accessed at [https://www.theguardian.com/technology/2021/nov/05/twitter-says-any-move-by-australia-to-ban-anonymous-accounts-would-not-reduce-abuse?CMP=Share\\_iOSApp\\_Other&s=09](https://www.theguardian.com/technology/2021/nov/05/twitter-says-any-move-by-australia-to-ban-anonymous-accounts-would-not-reduce-abuse?CMP=Share_iOSApp_Other&s=09)

<sup>24</sup> OAIC, *Australian Privacy Principles*, available at: <https://www.oaic.gov.au/privacy/australian-privacy-principles>, see APP 3 and APP 11.

<sup>25</sup>OAIC, "Chapter 3: Privacy Safeguard 3 – Seeking to collect CDR data from CDR participants", accessed at <https://www.oaic.gov.au/consumer-data-right/cdr-privacy-safeguard-guidelines/chapter-3-privacy-safeguard-3-seeking-to-collect-cdr-data-from-cdr-participants/>

<sup>26</sup> Doffman, Zak (23/8/2019), "Ashley Madison Has Signed 30 Million Cheating Spouses. Again. Has Anything Changed?", accessed at <https://www.forbes.com/sites/zakdoffman/2019/08/23/ashley-madison-is-back-with-30-million-cheating-spouses-signed-since-the-hack/?sh=5aac67123878>

<sup>27</sup> Twitter (10/8/2021), "Combatting online racist abuse: an update following the Euros", accessed at [https://blog.twitter.com/en\\_gb/topics/company/2020/combating-online-racist-abuse-an-update-following-the-euros](https://blog.twitter.com/en_gb/topics/company/2020/combating-online-racist-abuse-an-update-following-the-euros)



collection of identity verification is China. Under a series of laws, passed in 2017, Internet users in China must provide national identity documents and real names on a wide range of digital services<sup>28</sup>.

- 14.12. In 2004, Korea introduced an Internet real name system which, in 2007, was expanded to websites with more than 100,000 visitors per day under Article 44(5) of the Act on Promotion of Information and Communication Network Utilization and Information Protection. Data breaches ensued, with the personal information of 35 million Koreans being compromised in 2011<sup>29</sup>. These cyber security issues and other campaigning from Koreans about the law led to the removal of these requirements in 2012<sup>30</sup>.
- 14.13. DIGI is concerned that the focus around these current draft laws have not drawn Australians' attention to the fundamental changes that they are proposing to how Australians use the Internet. Do consumers want to routinely provide their personal information, age data, and potentially identity verification documents when they are perusing websites that include user interaction?  
**Per Recommendation #6, we believe that further consultation needs to occur on the propensity of Australians to undertake age verification and additional personal information as a pre-requirement to use digital services.**

## 15. Algorithms and online harms

- 15.1. 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.
- 15.2. On relevant large digital platforms, artificial intelligence (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. For example, machine learning enables Google's understanding of the use of varied language in search, to ensure user queries yield relevant information<sup>31</sup>. Algorithms enable users to use Google maps to navigate to their destination, and power grammatical and spelling corrections to texts, emails and online documents.
- 15.3. 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 Q3 of 2021, approximately 95% of videos removed from YouTube were detected by automatic flagging. This also allows videos to be removed before they are widely viewed; 36.7% were removed before they receive a single view<sup>32</sup>. Similarly, Meta uses artificial intelligence to proactively detect harmful content before it is seen by

---

<sup>28</sup> Samm Sacks & Paul Triolo (25/9/2017), *Shrinking Anonymity in Chinese Cyberspace - Lawfare*, accessed at <https://www.lawfareblog.com/shrinking-anonymity-chinese-cyberspace>

<sup>29</sup> Mills, Elinor (28/7/2011), "Report: Breach exposes data of 35 million S. Koreans", accessed at <https://www.cnet.com/tech/services-and-software/report-breach-e>

<sup>30</sup> Covington Inside Privacy, "Korea Strengthens Protection for 'Resident Registration Numbers' (RRNs): Leaks May Face a Fine of up to 0.5 Billion Korean Won", accessed at <https://www.insideprivacy.com/international/korea-strengthens-protection-for-resident-registration-numbers-rrns-leaks-may-face-a-fine-of-up-to-0/>

<sup>31</sup> Google AI Blog (2019), Understanding searches better than ever before, accessed at <https://www.blog.google/products/search/search-language-understanding-bert/>

<sup>32</sup> YouTube (2019), YouTube Community Guidelines enforcement, accessed at <https://transparencyreport.google.com/youtube-policy/removals>



- users. For content such as child exploitation material, terrorist content, violent and graphic content, and fake accounts, Meta proactively detects more than 99% of this content before a user needs to report it<sup>33</sup>.
- 15.4. The use of algorithms to promote online safety is consistent with the Government's expectations of industry. For example, the draft BOSE determination, that will come into effect with the OSA on January 23, 2022, identifies the detection of material and activity as a reasonable step service provider can take to ensure end users are safe<sup>34</sup>. 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<sup>35</sup>."
  - 15.5. Machine learning can also inform incerpation efforts in relation to online abuse. For example, Twitter uses behavioural signals to identify end-users who target other end-users with abuse or harassment and limits the visibility of their tweets. Facebook uses machine learning and a range of signals to identify posts from people who might be at risk of suicide, such as phrases in posts and concerned comments from friends and family, which involves a complex exercise in analysing human nuance, including analysis of the text in the post and the comments under the post. Once a cry for help is identified, Facebook may present the person with support options, including resources for help, help-line phone numbers, and ways to connect with loved ones.
  - 15.6. This speaks to a larger point that algorithms 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.
  - 15.7. Finally, AI can also be used to ensure the personalisation and relevance of information that a digital platform user sees. Often algorithms are dictated by the user's choices on which other accounts they choose to follow, and choices can be exercised with regard to hiding advertising.
  - 15.8. Platforms have been working to optimise their algorithms to prioritise content from authoritative sources, as well as reducing recommendations of "borderline" content that comes close to violating their policies or spreads harmful misinformation. For example, its May 2021 transparency report provided to DIGI as a signatory of the ACPDM, YouTube indicated that it has been amending its recommendation systems to reduce spread of so-called "borderline" content that toes the acceptable line, and has seen 70 per cent reduction toward its goal of getting it below 0.5% of all content viewed.
  - 15.9. DIGI is supportive of regulatory approaches to mitigate against defined harms, rather than specific technologies such as algorithms or AI. A harms-based approach reflects that the majority of potential problems associated with AI lie in the contextual application of the technology in a variety of sectors, and we caution against recommendations for regulatory or centralised bodies

<sup>33</sup>Meta (2021), "Community Standards Enforcement Report Q3 2021", accessed at <https://transparency.fb.com/data/community-standards-enforcement/>

<sup>34</sup> 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>

<sup>35</sup> 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>



focused solely on reviewing the technology of AI itself. Requirements to disclose specific technical details of the way in which algorithms operate, such as detailed information on the signals and predictions used, would not provide meaningful transparency to people and may serve to enable third parties to more easily game the system. For example, in the case of algorithms that are used to detect and remove harmful content, making them public would allow bad actors to manipulate posts to evade algorithm changes. In the immediate aftermath of the terrorist attacks in Christchurch in March 2019, platforms reported an unprecedented number of people actively manipulating the livestreamed footage of the attacks to avoid detection by algorithms.

## 16. Public figures and private citizens

- 16.1. DIGI recognises that the experiences of public figures, politicians and those in the public eye will differ on social media to the experience of private citizens. The majority of interactions that private citizens undertake on social media are with friends, family and people in communities within which they belong. By contrast, public figures and people in authority will have a higher volume of online interaction with people they do not know who are engaging with their contributions to public life. It is true that those interactions can cross a threshold of harm; relevant DIGI members apply their cyberbullying policies to public figures, and these do not allow credible threats to personal safety.
- 16.2. Another unique feature of Australia is “the absence of an explicit constitutional or statutory federal right to free speech”<sup>36</sup>. Such protections are not associated with the US alone, but exist in other comparable democracies; the UK has an enshrined right to freedom of expression under its Human Rights Act 1998 (UK), and Canada has constitutional entrenchment of the right free expression in the Canadian Charter of Rights and Freedoms (1982)<sup>37</sup>.
- 16.3. This contributes to the challenges faced by digital platforms in approximating Australians’ expectations with regard to freedom of expression and freedom of political communication. Platforms work from estimations of Australians’ expectations to be able to criticise those in the public eye because of their wide sphere of influence; that critique will sometimes be disrespectful, offensive or in poor taste, and may not meet the threshold for removal under platform policies. Decisions in this area will, by their very nature, never please all parties involved and there is a margin for error in the application of such policies in relation to public figures and, as with most online harms, platforms offer appeals mechanisms for when that occurs.
- 16.4. Where the content is considered to be defamatory, we expect that public figures and private individuals alike will have improved recourse once the the Council of Attorneys-General Defamation Working Party on the Review of Model Defamation Provisions (MDPs) completes its “Stage 2” process to ensure these provisions are fit for a digital age. DIGI is supportive of modernising these provisions to offer better solutions for Internet users, online intermediaries and complainants with regard to defamation.

<sup>36</sup> Gelber, Katharine (2019) “The precarious protection of free speech in Australia: the Banerji case” in Australian Journal of Human Rights.

<sup>37</sup> NSW Government (2021), *Attorneys-General Review of Model Defamation Provisions – Stage 2 (2021), Discussion Paper*, accessed at <https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/discussion-paper-stage-2.pdf>



- 16.5. Additionally, when the adult cyberbullying scheme comes into effect with the OSA on January 23, the regulatory guidance does not preclude politicians or public figures from seeking the eSafety Commissioner to issue a service provider with a removal notice that would require the removal of that content within 24 hours<sup>38</sup>.
- 16.6. **Per Recommendation #2, where victims of cyberbullying in particular are experiencing bullying on multiple platforms, in order to prevent the need for victims of abuse to have to report multiple pieces of content from a single perpetrator to multiple platforms, we encourage further strategic emphasis on the issuance of end-user notices in the Office's regulatory approach.**
- 16.7. Unfortunately, a legal gap persists in relation to hate speech which is not fully covered under the OSA's adult cyberbullying scheme for which the regulatory guidance states:
- A 'particular' Australian adult means one specific person, not a broad range or group of people. For example, racist abuse targeting a group rather than an individual, such as a post that says all people of a certain background 'should be wiped out' would not be adult cyber abuse for the purposes of this scheme because it is directed at a group rather than a specific person.*
- 16.8. While DIGI's relevant members all have long-standing platform policies prohibiting hate speech (regardless of whether the speech relates to public figures or private individuals), there is no recourse nor a comprehensive standard articulated in Australian law to guide the application of these policies across all digital services. **Per Recommendation #3, DIGI encourages the Australian Government to develop a clearer legislative framework that defines hate speech to help relevant stakeholders, including all digital platforms, to better report, review and remove content that meets a defined Australian legal threshold.**

## Section 3: Online harms on social media and online platforms

This section provides a high-level overview of a range of online harms, and summarises industry and regulatory approaches to each.

### 17. Overview of online harms, platform & regulatory responses

- 17.1. In this submission, DIGI seeks to advance the conversation about online harms on social media and online platforms, surface challenges and identify action areas. Working from a shared understanding of this extensive and evolving landscape is crucially important in building the Committee's understanding of online harms, and in identifying where there may be gaps that the Committee may wish to explore.
- 17.2. To that end, Table 1 below attempts to provide a high-level overview of the range of harms that Australians may experience on online platforms and the trends in industry responses to those.

---

<sup>38</sup> Office of the eSafety Commissioner, "Adult Cyber Abuse Scheme Regulatory Guidance", accessed at <https://www.esafety.gov.au/sites/default/files/2021-12/ACA%20Scheme%20Regulatory%20Guidance%20%20FINAL.pdf>



Table 1 also maps the current or forthcoming Australian Government regulation aimed at addressing the specific online harm.

- 17.3. 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. Where any of DIGI's members have provided their own submissions to this inquiry, we encourage the Committee to evaluate those for the relevant detail; Table 1 is simply a brief summary in order to provide an indication of the industry approach, and is by no means comprehensive.
- 17.4. Table 1 indicates the extent of platform and regulatory responses to online harms; from DIGI's vantage point, our experience is not that social media is "the Wild West" or that the Internet is an "unregulated space". The narrative and public messaging we use about online safety is important so as to encourage people that they have agency that they *can* take control of their online experience, availing of both the tools provided by digital platforms and the safety nets provided by the Australian Government.
- 17.5. Table 1 serves to illustrate the wide range of laws and initiatives underway pertaining to online safety, and demonstrates the complexity of the regulatory environment. **Per Recommendation #1, DIGI was very supportive of efforts to streamline legislation pertaining to online safety in order to aid clarity and compliance across the broad scope of the digital industry.**



**Table 1: Overview of online harms, platform & regulatory responses**

Online harm description	Trends in platform responses	Australian regulatory responses
<p>18. Cyberbullying material directed at an Australian child</p>	<p>18.1. 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.</p> <p>18.2. 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.</p> <p>18.3. This enforcement infrastructure is often complemented with proactive technology detection that detects problematic content and flags it for human review.</p> <p>18.4. 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.</p> <p>18.5. 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.</p>	<p>18.6. The Enhancing Online Safety Act 2015 (EOSA) allows Australian minors who are the target of cyberbullying material, and those representing them, to 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 48 hours. The Online Safety Act (OSA), which enters into force on January 23, 2022, reduces the timeframes that a social media service must respond to 24 hours.</p> <p>18.7. The draft Basic Online Safety Expectations (BOSE), when finalised, will come into effect with the OSA on January 23, 2022, and 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.</p> <p>18.8. The EOSA and OSA children's cyber bullying schemes enable 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. To date, DIGI understands that no such end-user notices relating to cyberbullying have ever been issued. <b>Per Recommendation #2, we believe that end user notices would help stem instigators of cyberbullying behaviour from using multiple platforms.</b></p>



		18.9. 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.
19. Cyberbullying material targeted at an Australian adult	<p>19.1. All of the measures outlined above from 18.1 to 18.4 (policies, tools, enforcement teams and technology) apply to the approach to cyberbullying material targeted at an Australian adult.</p> <p>19.2. 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.</p>	<p>19.3. The OSA, which will come into force on January 23, 2022, 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.</p> <p>19.4. The BOSE and Section 474.17(1) of the Criminal Code 1995 (Cth) detailed above in 2.7 and 2.8 also apply to adult cyberbullying.</p> <p>19.5. In relation to end users, the Australian Government made an election commitment May 5, 2019 to increase maximum penalties for end-users who use a carriage service to menace, harass or cause offence to five years of imprisonment<sup>39</sup>.</p>
20. Hate speech	<p>20.1. 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.</p> <p>20.2. 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.</p> <p>20.3. All of the measures outlined above from 18.1 to 18.4</p>	<p>20.5. 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.</p> <p>20.6. Australia continues to adopt a narrow approach to hate speech under anti- discrimination laws that are aimed at protecting individuals rather than groups based on their protected characteristics.</p> <p><b>20.7. Per Recommendation #3, DIGI has and continues to encourage the Australian Government to develop a clearer legislative framework that defines hate speech to assist</b></p>

<sup>39</sup> See media release: Prime Minister The Hon Scott Morrison MP, Attorney-General The Hon Christian Porter, Senator The Hon Mitch Fifield Minister For Communications And The Arts, Joint Media Release (05/05/2019), "Keeping Australians Safe Online", accessed at <https://www.liberal.org.au/latest-news/2019/05/05/keeping-australians-safe-online>

See also transcript: Prime Minister The Hon. Scott Morrison MP (05/05/2019), Transcript Remarks, Campaign Rally Central Coast, accessed via CCH alerts, see quote: "But the other thing we're going to do for all Australians, is we're going to increase the penalties for those who have been found to be bullying people online, causing those injuries. You won't go to jail for three years you'll go to jail for five years."

	<p>(policies, tools, enforcement teams and technology) apply to the approach to hate speech.</p> <p>20.4. Industry policies and enforcement are complemented with a range of initiatives, partnerships and social programs aimed at preventing and addressing hate speech.</p>	<p><b>enforcement agencies and prosecutors<sup>40</sup>. This will also serve to help relevant stakeholders, including digital platforms, to better report, review and remove content that meets a defined Australian legal threshold.</b></p>
<p>21. Defamation</p>	<p>21.1. Relevant DIGI members have policies that restrict the usage of their services for the defamation of others.</p> <p>21.2. 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 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.</p>	<p>21.3. 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 that existed prior to 2005. These provisions were not written for a digital age, and the Council of Attorneys-General Defamation Working Party on the Review of Model Defamation Provisions (MDPs), with a "Stage 2" process currently well underway to ensure these provisions are fit for a digital age. DIGI is supportive of modernising these provisions to offer better solutions for Internet users and online intermediaries with regard to defamation.</p> <p>21.4. From recent engagement with this defamation law reform process, DIGI understands that the NSW Law Reform Commission is considering a complaints notice process, debating using Section 5 of the UK 2013 Defamation Act as a starting point.</p> <p>21.5. In addition, on December 1, 2021, the Australian Government released the draft Social Media (Anti-trolling) Bill 2021 (ATB) that aims to aid the preliminary discovery process of potential defamation claims. DIGI intends to provide a separate submission on the ATB which we can provide to the committee once complete, if it is of interest.</p>

<sup>40</sup> DIGI, Submission to Department of Communications on the Online Safety Charter (14/04/09), accessed at <https://digi.org.au/advocacy/#:~:text=Online%20Safety%20Charter%20%7C%20Submission%20to%20Department%20of%20Communications>

<p>22. Misinformation and disinformation</p>	<p>22.1. Relevant DIGI members have policies and processes to remove or otherwise address the spread and scale of harmful misinformation and disinformation online. As with other policy areas described above, these policies are enforced through a combination of human review, proactive machine learning technology and enforcement teams.</p> <p>22.2. To provide a public, consistent and transparent framework for addressing the harm of mis- and disinformation to Australians, in February 2021, DIGI launched the <i>Australian Code of Practice on Disinformation and Misinformation</i> (ACPDM).</p> <p>22.3. 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:</p> <p>22.3.1.1. Publishing and implementing policies on their approach.</p> <p>22.3.1.2. Providing a way for their users to report content that may violate those policies.</p> <p>22.3.1.3. Implementing a range of scalable measures that reduce its spread and visibility online.</p> <p>22.4. 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 <a href="https://digi.org.au">digi.org.au</a>.</p> <p>22.5. The code contains opt-in commitments that have been widely adopted that entail (#2) Addressing disinformation in paid content. (#3) Addressing fake bots and accounts. (#4) Transparency about source of</p>	<p>22.8. DIGI developed the ACPDM in response to Australian Government policy announced in December 2019: <i>"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 for any further reform in 2021."</i><sup>42</sup></p> <p>22.9. DIGI understands that the ACMA provided their report on the effectiveness of the code to the Government on June 30, 2021, per the timeline requested by the Government.</p> <p>22.10. In the absence of seeing this report, and in advance of the review, over the course of September-November 2021, DIGI has made a recommendation to the Australian Government (via the ACMA, Minister's Office and the Department of Communications) for how the ACPDM can be strengthened, for which we are awaiting an outcome. DIGI has presented an identified gap in the governance arrangements outlined should disagreements arise between the complainants of the ACPDM and the Complaints Sub-Committee, and has proposed that the ACMA provide this appeals role. <b>Per Recommendation #4, we believe an appeals process operated by the ACMA will provide an important safety net for consumers in relation to the ACPDM.</b></p> <p>22.11. Separate to the ACPDM, DIGI has been working with representatives from the Electoral Council of Australia and New Zealand (ECANZ) on the development of escalation</p>
--	--	---

<sup>42</sup> Australian Government (11/12/2019), "Regulating in the Digital Age: Government Response and Implementation Roadmap", accessed at <https://treasury.gov.au/publication/p2019-41708>

	<p>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.</p> <p>22.6. 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.</p> <p>22.7. 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<sup>41</sup>.</p>	<p>processes during the forthcoming federal election.</p>
<p>23. Pro-terror material and the incitement of violence</p>	<p>23.1. DIGI members comply with Australian law and swiftly remove content that violates it, across a range of subject matter areas, including pro-terror content. They also work to report such content to law enforcement, where appropriate.</p> <p>23.2. Their policies prohibiting illegal pro-terror content form part of broader policies that prohibit the incitement or glorification of violence, and they rapidly remove content that may result in the credible risk of physical harm or direct threats to public safety.</p> <p>23.3. These policies are enforced through reporting tools, where end-users can escalate policy-violating content, and often through machine learning technology that proactively identifies potentially problematic content before many people have consumed it, both of which</p>	<p>23.7. The Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (AVM Act), passed in April 2019 requires content, internet and hosting providers to, within a reasonable time, report to the Australian Federal Police abhorrent violent conduct that is happening in Australia and accessible through their services, or hosted on their services.</p> <p>23.8. Additionally, the AVM Act requires the expeditious removal of abhorrent violent material, and provides the eSafety Commissioner the power to notify service providers that abhorrent violent material is available on their services. These notices create a presumption that the provider is aware of the material and puts providers on notice that such material should be removed<sup>45</sup>.</p> <p>23.9. The OSA also includes blocking notices for Internet Service</p>

<sup>41</sup> 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/>

<sup>45</sup> Attorney-General's Department, *Abhorrent violent material*, accessed at <https://www.ag.gov.au/crime/abhorrent-violent-material>

	<p>generally trigger a human review.</p> <p>23.4. With regard to pro-terror content specifically, several relevant DIGI members created a shared industry database of unique digital fingerprints, known as “hashes”, of known violent terrorist imagery or terrorist recruitment videos that had been removed from their services. Today, that database is used by thirteen companies that are members of the Global Internet Forum to Counter Terrorism (GIFCT). Companies rapidly used this database within hours of the Christchurch terrorist attacks adding over a thousand visually-distinct videos related to the attack to it. Crucially, these hashes were shared with smaller businesses to help stop the proliferation of this content on platforms that may not otherwise have the technology and resourcing of larger companies.</p> <p>23.5. This hash database is one example of industry collaboration that is occurring through the Global Internet Forum to Counter Terrorism (GIFCT), an NGO founded by several DIGI members that aims to (i) build shared technology to prevent and disrupt the spread of terrorist content online (ii) conduct and funding research by international experts, and (iii) share information and best practices with businesses of all sizes to assist them in managing this content on their platforms. Since 2017, GIFCT’s membership has expanded beyond the founding companies, and it has become an independent organisation.</p> <p>23.6. As one of several of its workstreams, the GIFCT has developed The Content Incident Protocol (CIP) to respond to emerging and active terrorist events, and assess any potential online content produced and disseminated by those involved in the planning or conducting of the attack. When the GIFCT declares the CIP is in force, all hashes of an attacker’s content</p>	<p>Providers for abhorrent violent conduct, alongside requirements for the takedown of other prohibited material detailed elsewhere in Table 1. <b>Per Recommendation #1, For clarity and to aid compliance across the breadth of in-scope companies, we would recommend the AVM Act be incorporated into a consolidated Online Safety Act.</b></p> <p>23.10. Furthermore, under the OSA, industry associations including DIGI have been asked to develop the new mandatory codes of practice to regulate all online services and websites available in Australia. These OSA codes must be developed in time to be registered by the Office of the eSafety Commissioner in July 2022<sup>46</sup>. On September 29, 2021, the Office of the eSafety Commissioner released a position paper<sup>47</sup> outlining expectations for the OSA codes.</p> <p>23.11. In subject matter, the OSA codes will relate to “Class 1” and “Class 2” materials under Australia’s classification code. <u>The list of Class 1 materials includes pro-terror content.</u> While there are specific requirements outlined in the position paper, at a high level, the OSA codes will contain commitments from industry to minimise the risk of harm to all Australian end-users due to the accessibility of Class 1 materials online. A code pertaining Class 1 material must be registered by the Commissioner around July 2022.</p> <p>23.12. <b>Per Recommendation #3, DIGI recommends the Australian Government provide further legal clarity by reviewing the protocol for listing terrorist organisations in response to the growing threat from the far right and consider whether new organisations should be added. This might be similar to the FBI list of Foreign Terrorist Organisations and the UK’s list of proscribed terrorist groups.</b></p>
--	--	---

<sup>46</sup> Online Safety Act 2021, see Part 9, Division 7, accessed at: <https://www.legislation.gov.au/Details/C2021A00076>

<sup>47</sup> Office of the eSafety Commissioner, *Development of industry codes under the Online Safety Act: Position Paper*, accessed at <https://www.esafety.gov.au/about-us/consultation-cooperation/industry-codes-position-paper>

	<p>are shared among the GIFCT's members, and a stream of communication is established between them. The first CIP was activated on October 9 2019, following the shooting in Halle, Germany<sup>43</sup>. In the wake of this shooting, the UN organisation Tech Against Terrorism confirmed<sup>44</sup> that measures taken by mainstream digital platforms resulted in a reduction in the virality of the livestreamed footage from Halle and observed that the footage was proliferating in smaller, less moderated forums.</p>	
<p>24. Child sexual abuse material (CSAM)</p>	<p>24.1. DIGI members have zero tolerance for CSAM. They have strict policies against child exploitation and the sexualisation of children. These policies are enforced through human review teams who undergo extensive training on the appropriate protocols for the handling of CSAM material, often with machine learning and other technology that surfaces content for review.</p> <p>24.2. When CSAM is detected it is removed and reported, DIGI members report to the National Center for Missing &amp; 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.</p> <p>24.3. 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.</p>	<p>24.5. The Online Content Scheme (Schedules 5 and 7 of the Broadcasting Services Act 1992) enables the eSafety Commissioner to investigate and take action on complaints about prohibited online content such as child sexual abuse material (CSAM).</p> <p>24.6. The EOsa and the Osa include 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.</p> <p>24.7. The AVm Act, detailed above, covers CSAM depicting rape or torture, which has been the subject of 98% of notices served under the Act<sup>48</sup>.</p> <p>24.8. Furthermore, the Osa Codes (detailed earlier in this submission and above in 23.10 and 23.11), relate to "Class 1" and "Class 2" materials under Australia's classification code. <u>The list of Class 1 materials includes CSAM</u>, and the code pertaining to Class 1 material must be registered by the Commissioner around July 2022.</p> <p>24.9. The draft BOSE, which when finalised will come into force with the Osa on January 23 2022, contains a specific expectation in Section 8 that service providers that use</p>

<sup>43</sup> Global Internet Forum to Counter Terrorism (GIFCT) website, accessed at <https://gifct.org/about/>

<sup>44</sup> Tech Against Terrorism (2019), "Analysis: What can we learn from the online response to the Halle terrorist attack?", accessed at <https://www.techagainstterrorism.org/2019/10/15/analysis-what-can-we-learn-from-the-online-response-to-the-halle-terrorist-attack/>

<sup>48</sup> Parliamentary Joint Committee on Law Enforcement, *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019*, report released December 2021, accessed via CCH political alerts.



	<p>24.4. 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).</p>	<p>encryption with their services will implement processes to detect and address material or activity on the service that is or may be unlawful or harmful. This implies that providers of encrypted services have special obligations to go further in scrutinising users' online interactions than providers of services that are more public.</p>
<p>25. Non-consensual sharing of intimate imagery</p>	<p>25.1. DIGI members have strict policies that do not allow the sharing of non-consensual intimate images, and work to rapidly remove these.</p> <p>25.2. These policies form part of broader policies to remove content that promotes sexual violence, sexual assault or sexual exploitation.</p> <p>25.3. As with other policy areas described above, these policies are enforced through a combination of human review, proactive machine learning technology and enforcement teams.</p> <p>25.4. 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.</p>	<p>25.5. The OSA, which will come into force on January 23, 2022, 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.</p> <p>25.6. The draft Basic Online Safety Expectations (BOSE), when finalised, will come into effect with the OSA on January 23 and 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 non-consensual intimate images are available, and to make reports about the provider's related activities available to the Commissioner.</p>
<p>26. Minors' access to pornography and other age-inappropriate content</p>	<p>26.1. 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 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.</p> <p>26.2. These policies are also reflected in members' advertising policies. For example, Google Search does</p>	<p>26.5. The forthcoming OSA codes, to be registered by the Office of the eSafety Commissioner in 2022, are expected to cover in scope the tools available to parents to manage and oversee their children's experiences online. DIGI and the Communications Alliance, supported by a steering group of other industry associations, are developing the new mandatory codes of practice to regulate all online services and websites available in Australia, which will be registered by the Office of the eSafety Commissioner in 2022. In subject matter, the codes will relate to "Class 1" and "Class 2" materials under Australia's classification code.</p>



	<p>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.</p> <p>26.3. 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.</p> <p>26.4. 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,</p>	<p>26.6. Class 2 materials include other online pornography, X18+ and R18+ content, and material which includes high-impact sex, nudity, violence, drug use, language and themes; 'Themes' includes social issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism. The OSA Codes will contain commitments from industry to minimise the risk of harm to Australian minors due to the accessibility of Class 2 materials online.</p> <p>26.7. Additionally, when the OSA enters into force on January 23, 2022, it will be accompanied by a new Restricted Access System (RAS) which requires that social media services, messaging services and websites limit access to certain age-inappropriate material through the implementation of an access control system. This will replace the 2014 Restricted Access System declaration<sup>49</sup>.</p> <p>26.8. In addition, the Office of the eSafety Commissioner is currently conducting a roadmap on age verification (AV Roadmap) – that was a result from Government's parliamentary inquiry into age verification for online wagering and online pornography. DIGI has engaged in consultations for the AV roadmap, and we understand that consultations are continuing and that the AV roadmap will be presented to the Government in 2022<sup>50</sup></p>
--	--	---

<sup>49</sup> Office of the eSafety Commissioner (2021), *Restricted Access System Declaration Online Safety Act 2021 Discussion Paper August 2021*, accessed at [https://www.esafety.gov.au/sites/default/files/2021-08/OSA%20-%20Restricted%20Access%20System%20discussion%20paper\\_0.pdf](https://www.esafety.gov.au/sites/default/files/2021-08/OSA%20-%20Restricted%20Access%20System%20discussion%20paper_0.pdf)

<sup>50</sup> Office of the eSafety Commissioner (16/8/21), *Consultations begin on age verification roadmap*, accessed at <https://www.esafety.gov.au/newsroom/media-releases/consultations-begin-on-age-verification-roadmap>

	<p>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.</p>	
<p>27. Advocacy of suicide and self-harm</p>	<p>27.1. 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.</p> <p>27.2. Relevant larger platforms partner with mental health organisations in Australia to produce or promote a range of training and other support resources.</p> <p>27.3. Such policies and partnerships also extend to material that glorify eating disorders such as anorexia nervosa, and bulimia.</p>	<p>27.4. Australia was the first country to criminalise pro-suicide websites in 2006 through the Criminal Code Amendment (Suicide Related Material Offences) Act 2005.</p> <p>27.5. It is possible that the aforementioned OSA codes to be registered in 2022 pertaining to Class 2 content cover such content in scope. As noted, Class 2 content has been defined as including “themes” that include “social issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism.”</p>
<p>28. Advertising of illegal and potentially harmful goods and services</p>	<p>28.1. 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.</p> <p>28.2. These policies include the prohibition of deceptive, misleading, or harmful business propositions, including restrictions on misleading, false, or</p>	<p>28.7. 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 the State and Territory consumer protection agencies.</p> <p>28.8. 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</p>

	<p>unsubstantiated claims during the promotion of a product or service.</p> <p>28.3. 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.</p> <p>28.4. Furthermore, there are restrictions on discrimination in the targeting of advertising to prevent discriminate against legally protected categories of customers.</p> <p>28.5. Members work hard to ensure that age-regulated advertising content, such as those for alcohol, are not served to minors.</p> <p>28.6. 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.</p>	<p>content on online content services in conjunction with live coverage of a sporting event.</p> <p>28.9. There are state and federal electoral laws that apply to digital content. As noted, DIGI has been working with representatives from the Electoral Council of Australia and New Zealand (ECANZ) on escalation processes with platforms for the upcoming federal election.</p>
<p>29. Scams, spam and deceptive conduct</p>	<p>29.1. 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.</p> <p>29.2. 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.</p>	<p>29.3. As noted in 12.7, Australian Consumer Law applies to digital platforms, and has prohibitions on false and misleading content</p> <p>29.4. The Australian Competition and Consumer Commission (ACCC)'s Scamwatch program enables consumers to complaint 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 educative functions.<sup>51</sup></p>

<sup>51</sup> NSW Fair Trading, "Scams and cybercrime", accessed at <https://www.fairtrading.nsw.gov.au/buying-products-and-services/scams>

<p>30. Privacy intrusion, hacking &amp; threats to cyber security</p>	<p>30.1. 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.</p> <p>30.2. Where applicable, they apply the strictest default privacy settings for minors; for example, ensuring that location-sharing is always off by default.</p> <p>30.3. DIGI members all allow their users to destroy, de-identify, access and correct their personal information in accordance with the Australian Privacy Act 1988 and where relevant they apply the European Union's General Data Protection Regulation (GDPR) requirements in this area.</p> <p>30.4. 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.</p>	<p>30.5. 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.</p> <p>30.6. Additionally, the Government has released for consultation an exposure draft of the Privacy Legislation Amendment (Enhancing Online Privacy and Other Measures) Bill 2021 (OPB). The OPB applies to social media services, large online platforms and data brokerage services. DIGI fully supports the intention of the Bill to protect the privacy of minors online. <b>Per Recommendation #5, we believe that specific privacy protections for minors should be expanded upon within the Privacy Act through the aforementioned Privacy Act Review to standardise these protections economy wide.</b> DIGI's primary concern with the Bill relates to its proposals for age verification, explored in Section 2 of this submission, which we believe carries privacy and cyber security risks.</p> <p>30.7. In relation to cyber security, the Department of Home Affairs is currently advancing work on Australia's cyber security regulations and incentives. It recently released options by way of a discussion paper titled <i>Strengthening Australia's cyber security regulations and incentives</i>. DIGI has engaged with this reform process and looks forward to continuing to engage in 2022.</p>
---	---	--



## 31. Shared challenges in addressing online harms

- 31.1. DIGI recognises that there are instances where more progress can be made. As mentioned, online safety and the management of online harms can never be a “set and forget” exercise. It requires continued and increasing investment through teams and technology, and research into emerging patterns of abuse and evolving community standards to inform continual iteration and improvement.
- 31.2. We have to place this dialogue in the context of the volume of content that is uploaded to digital platforms; for example, every minute, 500 hours of video are uploaded to YouTube and approximately 350 000 tweets are sent. While there is significant and increasing investment in online harm reduction, the scale of this challenge in a world where consumers expect to be able to instantaneously share content online cannot be underestimated.
- 31.3. 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 discussions of this issue, terms such as “fake news” are used to attack opponents who hold different views. In light of differing views, DIGI focused its definition of disinformation and misinformation on that which crosses a threshold of harm. We recognise that for 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.
- 31.4. Each of the online harms summarised in Table 1 mirrors a distinct and complex social or economic policy issue that manifests online. While DIGI and its 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.
- 31.5. 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<sup>52</sup>. Clear and accurate information from Governments is needed to build community understanding of issues

---

<sup>52</sup> UTS, *Discussion Paper on an Australian Voluntary Code of Practice for Disinformation*

Prepared for DIGI by UTS Centre for Media Transition, accessed at

<https://digi.org.au/wp-content/uploads/2021/02/Discussion-Paper-ACPDm-FINAL-PDF-Updated-Feb-2021.pdf>, p. 16.



prone to misinformation, and we need to improve digital literacy in the community. 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.

## 32. Conclusion

- 32.1. DIGI and its members believe that the digital industry has an immense responsibility to address online harm, and that the Australian Government has an important role to play in standardising protections, encouraging accountability and providing safety nets for consumers. 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.
- 32.2. 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 combatting online harms in 2022 and beyond.