February 13, 2021

Director, Digital Platforms and Online Safety
Department of Infrastructure, Transport, Regional Development and Communications
By email: OnlineSafety@infrastructure.gov.au

Dear Director,

Thank you for the opportunity to engage with the Australian Government about its proposed Consultation on a Bill for a new Online Safety Act ("the Bill"), released on December 23, 2020.

By way of background, DIGI is a non-profit industry association that advocates for the interests of the digital industry in Australia, with Google, Facebook, Twitter and Verizon Media as its founding members. DIGI also has an associate membership program and our other members include Redbubble, eBay, Change.org and GoFundMe. 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 founding members publish detailed information about their specific efforts in relation to online safety, including transparency reports and strict policies outlining restricted content and user behaviour on their platforms, which are regularly updated to ensure they reflect emerging patterns of abuse. They have heavily invested in reporting tools and content moderation teams to ensure illegal and policy-violating content is surfaced and promptly actioned, along with expedited processes and protocols for content that requires rapid response.

The industry has and continues to invest in technology to detect and prevent the dissemination of policy-violating content, including image hashing classifiers to report and identify child sexual exploitation material, a hash database of URLs directing to known terrorist content shared among companies, and machine learning algorithms that proactively identify potentially problematic content for human review. They work closely with the Australian Government, governments around the world and civil society to address a wide range of issues related to online safety; this includes extremely close ongoing collaboration and working relationships within the Office of the eSafety Commissioner.

All that is to say, DIGI shares the Government's strong commitment to online safety and our founding members have and continue to make major, longstanding investments in the safety of their users and the community. DIGI is supportive of efforts to streamline legislation pertaining to online safety under one consolidated Online Safety Act.

However, this submission raises concerns about the scope, transparency and implementation of the Bill, that we hope will be taken into account when the Act is finalised. Some of the many concerns raised in this submission include the broad range of digital products and services in scope under the Act, and the potential for the subjective thresholds of harm in the adult cyberbullying scheme to limit legitimate political expression. DIGI is also concerned about the high amount of discretion the Bill's proposals vest within the eSafety Commissioner's Office (the Office) without any procedural fairness nor transparency in relation to removal notices, service provider notifications, transparency report demands, industry standards and applications for the cessation of services to the Federal Court.

While we support the updating of online safety legislation to keep pace with changes in the online environment, it is important to ensure that reforms can achieve their objectives in a way that provides

an appropriate level of accountability for both the company activities that are regulated, and the regulatory body tasked with enforcing the relevant regulations. This balance is essential to providing a firm basis for evaluating the success of the reform program.

Thank you again for the opportunity to input on the Bill. Should you wish to discuss any of the representations made in this submission further, please do not hesitate to contact me.

Best regards,

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

Table of contents

1. Objects of the new Act

DIGI is supportive of the Government's objective of improving and promoting online safety for Australians. Our founding members have and continue to make major longstanding investments in the safety of their users and the community. We welcome the refinement of these objects from the objects proposed in the Online Safety Legislative Reform Discussion Paper ("the Discussion Paper"), released in December 2019. The previous object for the proposed Act set out in Section 3a of the Discussion Paper of "preventing online harms" is not achieved through the takedown of content alone, as the removal of content is a remedy to address harmful content rather than a means to prevent it from occurring.

2. Scope of services covered

DIGI is extremely concerned about the expansive scope of services within the digital industry covered under the Bill. Clarity and refinement of the scope of digital services that may be covered in the Bill is needed.

2.1 Enterprise services

The Bill covers "social media services", "electronic services", and "designated internet services", and each of the definitions of these terms contain references to "end users". However, the Bill does not provide a definition of "end users". It is unclear whether "end user" captures individuals within enterprises, thereby creating uncertainty for enterprise software and B2B digital services as to whether they are captured under the Bill.

2.2 Private messaging & email

In addition, the definition of "social media service" is also overbroad wherein it states "the sole or primary purpose of the service is to enable online social interaction between 2 or more end-users". This definition suggests that the Bill extends to services offering any personal messaging between two individuals. Therefore, in its current form, the Bill creates an obligation for those services that does not exist in relation to text messaging through SMS and MMS technology, nor does it provide a compelling rationale for this discrepancy. It is important to additionally note that most mainstream online private messaging and email services have a block function and/or report function that prevents unsolicited and otherwise unwanted messaging, that would encompass cyberbullying and other content types covered under the Bill.

This definition would affect a broad range of digital products and services. In general, we would define "private messaging" services to include email products (e.g. Gmail, Yahoo mail), including email products offered within enterprise services (e.g. Outlook), private messaging associated with user-generated content platforms (e.g. Facebook Messenger, Twitter Direct Message), those that use Internet connectivity for the transmission of messages (e.g. Whatsapp), including those that are associated with particular hardware (e.g. Apple iMessage) or enterprise services (e.g. Slack). As these examples illustrate, there is a broad set or services covered in the Bill as currently drafted.

Granting the Office takedown powers over content transmitted through such private messaging forums is not a response that will serve to deter or prevent perpetrators of abuse, and those that send illegal content. First of all, the removal of content from a takedown scheme is not a logical solution to address the concerns of the complainant; the result will simply be that they are no longer able to view the private messages in their mobile or desktop record of a private conversation. A more logical approach would be to prevent the perpetrator of the abuse from sending future messages, which would be achieved through platforms continuing to ensure there are blocking and/or reporting functionalities within the provision of a private messaging or email service.

A longer term deterrent would also be criminal penalties for the perpetrators of abuse, such as through measures like the Australian Government's election commitment on May 5, 2019 to increase maximum penalties for using a carriage service to menace, harass or cause offence. The Government also announced new offenses relating to dealings with child abuse material, grooming third parties using the post or a carriage service to procure children for sexual activity, and indecent communication to a child. These penalties, and the provision of in-platform blocking and reporting functions, are more effective ways to deter and address cyberbullying that occurs in private messaging; the reliance on the takedown of such content alone is not effective in achieving the Objects of the Act.

2.3 Scope and exclusions in other jurisdictions

The scope of the Bill should be refined, drawing on the recent experience of other jurisdictions. In the UK Government's response to the Online Harms White Paper, it acknowledged concerns about what it deemed "low-risk businesses" being captured in scope of the new framework, and decided to focus on developing a framework "designed to reduce the burden on UK business by focussing on the areas that present the greatest risk of harm"[1]. As a result, the UK Government created the following exemptions to its scheme:

> *"Business services. Online services which are used internally by organisations - such as intranets, customer relationship management systems, enterprise cloud storage, productivity tools and enterprise conferencing software - will be excluded from scope. The risk of harm on*

[1]UK Government, *Consultation outcome Online Harms White Paper: Full government response to the consultation Updated 15 December 2020*, available at https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#contents

these services is low, as the user base is limited and users tend to be verified and acting in a professional capacity. Organisations will already have policies in place for protecting users and managing disputes. Requiring them to comply with the legislation would be a disproportionate regulatory burden.

Online services managed by educational institutions, where those institutions are already subject to sufficient safeguarding duties or expectations. This includes platforms used by teachers, students, parents and alumni to communicate and collaborate. This is to avoid unnecessarily adding to any online safeguarding regulatory or inspection frameworks (or similar processes) already in place.

Email and telephony. Email communication, voice-only calls and SMS/MMS remain outside the scope of legislation. It is not clear what intermediary steps providers could be expected to take to tackle harm on these services before needing to resort to monitoring communications, so imposing a duty of care would be disproportionate."[2]

Furthermore, the German Netzwerkdurchsetzungsgesetz (NetzDG) law has refrained from including private messaging services in its scope. It is also limited to public "social networks" that have over two million users in Germany, and pertains only to content that is unlawful under the German criminal code:

Section 1 Scope

(1) This Act shall apply to telemedia service providers which, for profit-making purposes, operate internet platforms which are designed to enable users to share any content with other users or to make such content available to the public (social networks). Platforms offering journalistic or editorial content, the responsibility for which lies with the service provider itself, shall not constitute social networks within the meaning of this Act. The same shall apply to platforms which are designed to enable individual communication or the dissemination of specific content.

(2) The provider of a social network shall be exempt from the obligations stipulated in sections 2 and 3 if the social network has fewer than two million registered users in the Federal Republic of Germany.

(3) Unlawful content shall be content within the meaning of subsection (1) which fulfils the requirements of the offences described in sections 86, 86a, 89a, 91, 100a, 111, 126, 129 to 129b, 130, 131, 140, 166, 184b in connection with 184d, 185 to 187, 241 or 269 of the Criminal Code and which is not justified.[3]

Recommendations 1-5

1. A definition of "end user" should be included to specify that this is a member of the general public. Services in scope should be those of a business to consumer (B2C) nature, rather than those that are business-to-business (B2B).

2. Enterprise services should be excluded from the scope of the Bill. These include software as a service (SaaS) where companies licence software to provide to their employees, and the infrastructure and data are hosted in the service provider's data centre, usually using cloud-based computing. The use of such software is typically governed by the licencing organisation's human resources policies, which would almost certainly restrict the

[2] UK Government, *Consultation outcome Online Harms White Paper: Full government response to the consultation Updated 15 December 2020*, available at https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#contents

[3] German Law Archive, *Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)*, available at https://germanlawarchive.iuscomp.org/?p=1245

transmission of cyberbullying, image-based abuse and the distribution of Class 1 and Class 2 material under the Bill.

3. Private messaging services and email should be excluded from the cyberbullying takedown schemes. The Bill's scope should be restricted to public content sharing services that enable the upload of user-generated content. This may include user-generated content platforms, including blogs, social media and networks (e.g. Facebook, Twitter), discussion boards and photo, text and video sharing sites (e.g. YouTube). These are services where content can be viewed by the general public or other broad audiences, and where takedown can provide an effective remedy to victims of cyberbullying and other restricted content.

4. To provide remedy to victims of cyberbullying, image-based abuse and recipients of illegal and harmful content contained within private messaging and email, the Bill should alternatively:
    a. State in the Bill under the Basic Online Safety Expectation (BOSE) framework, and any further guidance released on this framework, that providers of private messaging and email services should provide a blocking function and/or reporting function that is available to users to restrict incoming messages from designated recipients.
    b. Focus on the Office's ability to issue end user notices to the perpetrators of abuse in private messaging. This is a far more scalable and targeted solution that would also serve to prevent abuse from occurring on multiple private messaging platforms.

5. As discussed above, cloud infrastructure and hosting services should not be subject to the takedown schemes involving removal notices. However, we understand that the Office may wish to reserve the right to capture cloud and hosting services as a matter of contingency, should there be technical and/or contractual issues preventing the removal of content on the service upon which it was originally posted. If this is the case, then there must be a hierarchy in the order by which the Office can serve notices to indicate that a removal notice to a hosting service provider is a matter of last resort, that can be served only after a removal notice is issued to an end user or a public content sharing service. This tiering would recognise that hosting service providers are the most removed the content that is the subject of the removal notice, and will therefore have the most blunt instruments in order to remove the content in question that will impact a wider set of end-users. In addition, the requirement to comply with takedown schemes within 24 hours will prevent the deployment of more targeted solutions to remove only the content in question, and thereby limit the number of affected end users.

3. Basic online safety expectations (BOSE)

DIGI's members share the Government's goal and expectation that technology companies and digital platforms should be proactive in ensuring online safety in their products and services. We welcome the clarification in the Bill as to how the BOSE framework differentiates from the eSafety Office's Safety by Design principles, which was previously unclear when the BOSE framework was first advanced in the Discussion Paper.

3.1 Demands for public statements

Further clarification is needed under Part 4 Division 2 Section 46, 1 g) h) and i) where it is stated that:

> *g) the expectation that, if the Commissioner, by written notice given to the provider of the service, requests the provider to give the Commissioner a statement that sets out the number of complaints made to the provider during a specified period (not shorter than 6 months) about breaches of the service's terms of use, the provider will comply with the request within 30 days after the notice of request is given;*
>
> *h) the expectation that, if the Commissioner, by written notice given to the provider of the service, requests the provider to give the Commissioner a statement that sets out, for each*

removal notice given to the provider during a specified period (not shorter than 6 months), how long it took the provider to comply with the removal notice, the provider will comply with the request within 30 days after the notice of request is given;

i) the expectation that, if the Commissioner, by written notice given to the provider of the service, requests the provider to give the Commissioner specified information relating to the measures taken by the provider to ensure that end-users are able to use the service in a safe manner, the provider will comply with the request within 30 days after the notice of request is given.

The criteria for which the Commissioner may make such demands for public statements to a provider is wholly unclear. Given the extremely high level of reporting burden to provide details of complaints received over a six month period, within 30 days after the notice of the request is given, there should be a consistent and publicly available rationale for the Office's request of these reports.

Recommendations 6-9

6. The Bill must specify a clear, consistent and transparent criteria for the requests for public statements under Section 46.

7. The criteria mentioned in Recommendation 6 must be on the basis of documented, systemic violations of the BOSE. Focusing this rationale on systemic violations of the BOSE will mitigate the reputational damage and compliance burden that will arise from being the subject of this report request for companies that are working in good faith to meet the BOSE.

8. There must be procedural fairness for the issuance of the reports, and a documented administrative process that the Commissioner will undertake prior to their issuance.

9. There must be a documented process included in the Bill for providers to challenge the reports for not meeting the specified criteria under Recommendation 6.

4. Cyberbullying scheme for children

DIGI shares the Government's strong commitment to protecting minors online. DIGI founding members employ a range of tools in this area including requiring minimum age requirements for account creation, age restrictions, strict policies that prohibit the cyberbullying of children, processes to swiftly address reports of violations of those restrictions, and an enforcement infrastructure comprised of proactive technology detection and human moderators. They also have tools to restrict the experience of minors online, and also invest in social programs aimed at minors and parents to promote safe experiences online.

4.1 Scope of services covered

DIGI notes that the Bill extends the existing scheme under the Enhancing Online Safety Act (EOSA) to more services. DIGI reiterates the concerns and related Recommendations 1-5 detailed earlier in Section 2 in relation to limiting the scope of services covered.

4.2 End user notices

We welcome the provision for the Commissioner to issue an end user notice to apologise or to refrain from posting cyber-bullying material targeted at the complainant in the future. This is a scalable solution, as it may serve to deter the end user from posting material on different providers' services.

We encourage the consideration of more behavioural and perpetrator level policy approaches. It is worth noting that the current 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. It is important that digital providers and the public have transparency into the decision-making processes of the Office, including an understanding of why powers are not exercised. This will assist efforts to evaluate the effectiveness of the current and proposed schemes, any gaps in those schemes, and whether there is a need for increased powers for the Commissioner as a result.

4.3 24 hour turnaround time

We seek clarifications and a justification in relation to the shortening of the turnaround time from 48 hours under EOSA to 24 hours for a provider to remove content in response to a removal notice under this and other takedown schemes included in the Bill. As noted in Section 4.2., it is important to understand the effectiveness and perceived gaps in the current scheme, and align the Bill to the assessment of those gaps. The Discussion Paper states: "The eSafety Commissioner has had great success in working with social media companies to remove material in very short time frames – even as short as 30 minutes"[4]. Therefore, the justification for the shortening of these timeframes is unclear.

DIGI's members promptly respond to all communications from the Office under the current EOSA schemes, including formal notices and informal communication, and have rapid response protocols in place with the eSafety Office such that these are usually actioned well within 24 hours. However, particularly given the scope of the services covered under the Bill and their varying sizes, we do see merit in clarifying the process and intermediary liability in exceptional cases where investigation and due diligence in relation to a complaint may necessitate more than 24 hours. Given the proposal to shorten time frames to 24 hours across all types of content covered under the proposed Act, these concerns will be further elaborated upon in the following section in relation to the cyberbullying scheme for adults, where there are more frequently factors that may necessitate a longer timeframe than content involving minors.

Recommendations 10-13

10. We recommend that the Government establish an independent review, separate to the BIll, as to why no end-user notices relating to cyberbullying have been issued to date by the Office of the eSafety Commissioner, under EOSA.

11. A proper process for a provider to respond in writing to a removal notice under the cyberbullying scheme for children must be included in the Bill. This process should provide the provider with an opportunity to outline the reasons why the content in question has or has not been removed and detail any contextual factors that have become apparent in their investigation of the issue.

12. This process should also clarify the intermediary liability status of the provider during the stages of the process outlined in Recommendation 11, pertaining to the cyberbullying scheme for children.

13. There must be a provision for the Commissioner to assess the provider's response, under Recommendation 11, in determining whether to re-issue the removal notice under the cyberbullying scheme for children.

[4] *Online Safety Legislation Reform - Discussion paper*, published December 11 2019, available at https://www.communications.gov.au/have-your-say/consultation-online-safety-reforms, p.1

5. Cyberbullying scheme for adults

We understand that this is a new scheme that allows for the removal of material that seriously harms Australian adults, with similar protections in the cyber-bullying scheme to adults, however with a higher threshold of 'harm' to reflect adults' higher levels of resilience. Every DIGI member has policies to restrict content and user behaviour on their platforms in relation to bullying, harassment or abuse that directly threatens another person. These policies are regularly updated to ensure they reflect emerging patterns of abuse. In addition, the industry has also heavily invested in reporting tools and content moderation teams to ensure policy-violating content is surfaced and promptly actioned. They also have expedited processes and protocols in place for urgent reports from law enforcement bodies, and for other content that requires rapid response. Reports of policy-violating and illegal content are reviewed and actioned by real people, who undergo extensive initial and ongoing training.

5.1 Scope of services covered

We repeat the same concerns and recommendations articulated above that the scope of services covered under the scheme is too broad, and should be limited to public content sharing services that enable the upload of user-generated content. We reiterate the Recommendations 1-5 detailed earlier in relation to limiting the scope of services covered.

5.2 48-hour threshold for removal notice issuance

We understand that one of the criteria that enables the Commissioner to issue a removal notice is if a complaint was made to the provider's service, and the content in question was not removed within 48 hours (see Section 88, 1, d). While most clear, prima facie examples of cyberbullying will be removed well within 24 hours on DIGI member platforms, there will be cases where the determination that content requires removal will not be immediately apparent and may necessitate further investigation, often with the claimants and content authors, and turnaround time for the decision about content removal may exceed 48 hours in such cases.

There will also be cases where, upon investigation of the claim, the provider determines that the content does not meet their adult cyberbullying policy standard for removal. 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. Under the current scheme, there is no procedure for a platform to provide this contextual information behind their determination not to remove content, and challenge the Commissioner's removal notice.

5.3 Service provider notifications

Under Section 93, we understand the Commissioner may notify the public where there were two or more occasions during the previous 12 months on which cyber-bullying material targeted at an Australian adult was provided on a service, and the material contravened that service's terms of use. Specifically, the Bill states:

> *(2) the Commissioner may, with the consent of the complainant, give the provider of the service a written notice that:*
> *(e) identifies the material; and*
> *(f) states the Commissioner is satisfied that the material is cyber-abuse material targeted at an Australian adult. If the Commissioner is satisfied that there were 2 or more occasions during the previous 12 months on which:*
> *(a) cyber-abuse material targeted at an Australian adult was provided on:*
> *(i) a social media service; or*
> *(ii) a relevant electronic service; or*

(iii) a designated internet service; ...

The criteria for the provision of these notifications is wholly unclear in the above description, and sets a low threshold for the Commissioner's issuance of these notifications. Taking into account the wide scope of services covered (as explored in Section 2 of this submission), it is almost certain that any of these services will have experienced "2 or more occasions during the previous 12 months" of cyberbullying content on their platforms. This stated threshold also does not take into account whether the content was actioned by the service. As a result, the Commissioner has grounds to issue service provider notifications to *any* social media service, relevant electronic service or designated internet service, inflicting significant reputational damage on that service, regardless of whether they are complying in good faith with the Bill and otherwise engaging in best practice. As a result, we would suggest the criteria be significantly modified and limited to public content sharing services that have systemic, documented violations of the BOSE framework and/or violations of the Commissioner's removal notices.

5.4 24 hour response time to removal notice

While most clear, prima facie examples of cyberbullying will be removed well within 24 hours on DIGI member platforms, there are often reasonable grounds for determinations that may take longer -- as noted elsewhere in this section of the submission. It is concerning that a justification has not been provided for why a provider must remove content in response to a removal notice within 24 hours.

In relation to the proposed timeframes of responding to notices, the Discussion Paper indicates that the 24-hour response time is "consistent with international practice for take-down of illegal and harmful content" -- however, this statement is incorrect. By contrast, the Netzwerkdurchsetzungsgesetz ("NetzDG") law in Germany relates only to "illegal" content by cross-referencing the German Criminal Code, as is evidenced from the excerpt of the law provided in Section 2.3 of this submission. Content that is not clearly illegal under NetzDG is subject to a seven day review period[5]. There is no subjective discretion for any regulatory body in Germany to demand content be removed, nor to demand that such content is removed within 24 hours on that basis.

In this regard, we would argue that the eSafety Commissioner should operate to uphold Australian law; it is not appropriate that the Commissioner would take a role in making subjective judgements about whether a company is upholding its Terms of Service beyond compliance with Australian law. This creates confusion, particularly as most platforms have extensive internal operational policy manuals to implement their own Terms of Service in a way that is bespoke to that company's service and its community of users. That is to say, a company is best placed to determine whether content violates its Terms of Service, and the regulator is best placed to determine whether content violates the law. The Commissioner may still alert platforms of content that they believe violates their Terms of Service and may warrant removal, but in situations where the content does not also violate the law, this should not be considered a legal directive under the Act, but rather a part of the cooperative and voluntary working relationship between the eSafety Office and the digital industry.

Furthermore, it is important to emphasise that NetzDG does not always require providers to remove content within 24 hours; the timeframe for the assessment whether or not a post or comment has to be deleted depends on how clearly the content violates any of the relevant criminal codes, defined under a pre-existing and separate part of German law to NetzDG. If the content clearly and obviously violates one of these criminal codes, the content has to be deleted within 24 hours of the user's complaint. If it is unclear if a German criminal code has been violated, the content has to be assessed more carefully within a 7 day period. If a thorough assessment of the content leads to the conclusion that it is illegal, it has to be deleted within seven days of the user's complaint.

By stark contrast to NetzDG, the Australian Bill provides:

- No discretion for a provider to assess the content that is the subject of the removal notice, under their own Terms of Service;

[5] German Law Archive, *Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG)*, available at https://germanlawarchive.iuscomp.org/?p=1245

- No discretion for a provider to assess the content that is the subject of the removal notice under the descriptions provided in the Bill;
- No mechanism for the provider to provide the context of their assessment to the Commissioner during or after the 24 hour removal notice period for the Commissioner to reconsider whether the issuance of the removal notice is appropriate.

The above points must be addressed in the final text of the Act.

5.5 Threshold of harm

We note that the threshold of harm to determine that material is cyber-abuse targeting an Australian adult is that:

> *b) "an ordinary reasonable person would conclude that it is likely that the material was intended to have an effect of causing serious harm to a particular Australian adult;*
> *c) an ordinary reasonable person in the position of the Australian adult would regard the material as being, in all the circumstances, menacing, harassing or offensive;*

Whereby "serious harm" is defined as:

> *serious harm means serious physical harm or serious harm to a person's mental health, whether temporary or permanent.*
>
> *serious harm to a person's mental health includes: (a) serious psychological harm; and (b) serious distress.*

The statutory test based on the conclusions of "an ordinary reasonable person" is extremely challenging to interpret. In addition, "serious harm" is a subjective concept, and would be entirely dependent on the judgement of the Commissioner. "Offensive" is also highly subjective, and may set a low threshold for content removal.

In practice, as mentioned, digital providers have far more granular considerations when assessing the cyberbullying of adults. For example, some of their considerations might include questions like: Does the content concern private positions vs. public opinions or actions? Does the content concern someone's professional role, employer or actions at work that may impact other people? Is the content about a public figure, person in authority, or private individual? The questions a platform may ask will necessarily differ based on the service, and provide important checks and balances for platforms to appropriately consider the freedom of expression implications of a takedown decision. The current subjective thresholds of "an ordinary reasonable person", "serious harm" and "offensive", taken either individually or together, may be used to silence political speech or other legitimate commentary.

On this point, Section 233 of the Bill indicates that:

> *This Act does not apply to the extent (if any) that it would infringe any constitutional doctrine of implied freedom of political communication.*

Yet there is nothing in the statutory test above that precludes the censorship of political communication. The definition of the cyberbullying of adults must be reconsidered in this context. Unlike the cyberbullying of children, where responsible digital providers err on the side of content removal in order to protect minors, there are greater implications for the potential silencing of legitimate expression that is in the public interest if the definition of adult cyberbullying is not developed in a highly considered manner, that applies objective and critical analysis to any decision to remove content.

5.6 Defamation overlap

The final Act must also give due consideration to the intersection between the cyberbullying of adults and online defamation, for which there is currently a national law reform review and a forthcoming "Stage 2" process that is specifically focused on defamation on digital platforms. This is because there can be a clear overlap between what is considered bullying and defamation: for example, a statement where one person calls another a fascist can be considered under the lens of both bullying and defamation. Definitions need to be clearly delineated to provide meaningful guidance to complainants and providers about their rights and responsibilities under the adult cyberbullying scheme, and the outcome of the Stage 2 defamation law reform process in relation to digital platforms.

Consistent with legal approaches globally to other illegal content -- such as the UK 2013 Defamation Act, which we understand is being examined in the context of Australia's own current defamation law reform currently underway -- the legal position of an online intermediary needs to be made abundantly clear during the time in which it is examining a takedown claim under any law pertaining to online content. If the Government elects to use a prescribed turnaround time as the measurement of compliance under any new Online Safety Act, it should also provide legal protection for organisations where there are legitimate circumstances that mean that reviewing and responding to the complaint may take longer.

Recommendations 14-19

14. The 48-hour threshold, expressed in Section 88 of the Bill, pertaining to the criteria for the Commissioner's issuance of a removal notice, must be extended.

15. A process for a provider to respond to a removal notice under the cyberbullying scheme for adults must be included in the Bill. This process should outline the reasons why the provider may have or have not removed the content in question, and any contextual factors that have become apparent in their investigation.

16. This process outlined in Recommendation 15 should also clarify the intermediary liability status of the provider during the stages of the process, pertaining to the cyberbullying scheme for children.

17. There must be a provision for the Commissioner to assess the provider's response, under Recommendation 15, in determining whether to re-issue the removal notice under the cyberbullying scheme for adults.

18. The threshold of harm should be reconsidered, and should also include counterbalancing protections for consideration of freedom of expression and political communication.

19. Clear delineation between the adult cyberbullying scheme and applicable defamation law must be provided in the Bill, so as to provide meaningful guidance to both providers and complainants.

6. Non-consensual sharing of intimate images

6.1 Removal notices

To a social media service, relevant electronic service or designated internet services

DIGI members have strict policies that do not allow the sharing of non-consensual intimate images. The response times for the removal of image-based abuse by major platforms once reported are extremely fast and well within the 24 hour proposal. Some platforms have also introduced preventative measures that use image hashing to prevent the spread of known image-based abuse images to prevent the reliance on user reporting. That said, and as discussed in relation to other

takedown schemes, codifying a 24 hour turnaround time into legislation is problematic in certain cases that require more complex technical solutions, investigation, and for smaller less resourced companies. The Bill should clarify the legal position of intermediaries in exceptional cases that may necessitate more than 24 hours for content to be removed.

To an end-user

We welcome the inclusion of end user removal notices. As noted, the issuance of end-user notices is an effective deterrent to perpetrators of image-based abuse, and will also serve to prevent the posting of non-consensual intimate images on various services. End-user notices and penalties should be well publicised through public communications campaigns in order to raise awareness of the criminal nature of image-based abuse, and to deter the occurrence of it.

To a hosting service provider

We note that the Bill extends the existing scheme under the Enhancing Online Safety Act (EOSA) to more services and we reiterate the concerns and reiterate Recommendations 1-5 in Section 2 of this submission.

Echoing these scope concerns, hosting service providers are the most removed from the content that is the subject of the removal notice, and will therefore have the most blunt instruments in order to remove the content in question that will impact a wider set of end-users. In addition, the requirement to comply with takedown schemes within 24 hours will prevent the deployment of more targeted solutions to remove only the content in question, and thereby limit the number of affected end users. We understand that the eSafety office may wish to reserve the right to capture cloud and hosting services as a matter of contingency, should there be issues preventing the removal of content on the service upon which it was originally posted. If this is the case, then there must be a hierarchy in the order by which the eSafety office can serve notices to indicate that a removal notice to a hosting service provider is a matter of last resort, that can be served only after a removal notice is issued to an end user or a public content sharing service.

6.2 Service provider notifications

Under Section 93, we understand the Commissioner may notify the public where there were two or more occasions during the previous 12 months on which cyber-bullying material targeted at an Australian adult was provided on a social media service, relevant electronic service or designated internet service, and the material contravened that service's terms of use.

As with the adult cyberbullying scheme, the criteria for the provision of these notifications is wholly unclear, and sets a low threshold for the Commissioner's issuance of these notifications. Taking into account the wide scope of services covered (as previously explored), it is highly likely that any of these services will have experienced "2 or more occasions during the previous 12 months" of image-based abuse on their platforms. This stated threshold does not take into account whether the content was actioned by the service. As a result, the Commissioner has grounds to issue service provider notifications to any social media service, relevant electronic service or designated internet service, inflicting significant reputational damage on that service, regardless of whether they are complying in good faith with the Bill and otherwise engaging in best practice. As a result, we would suggest the criteria be significantly modified and limited to public content sharing services that have systemic, documented violations of the BOSE framework and/or violations of the Commissioner's removal notices.

Recommendations 20-21

20. End-user notices and penalties should be well publicised through public communications campaigns in order to raise awareness of the criminal nature of image-based abuse, and to deter the occurrence of it.

21. The criteria for service provider notifications under the non-consensual sharing of images scheme should be significantly modified and limited to public content sharing services that

have systemic, documented violations of the BOSE framework and/or violations of the Commissioner's removal notices.

7. Online content scheme

7.1. Removal notices

RC and X18+ content violates most responsible digital platforms' Terms of Service; all DIGI members have strict content policies in relation to pornographic content, including child sexual exploitation material, and violent and graphic content. On social media and content platforms, there are prohibitions in their community guidelines on nudity, pornography and sexual explicit content including that which includes minors, as well as content that glorifies violence. On Google Search, sexual and violent terms are removed from auto-complete and pornographic results are demoted in ranking unless the user is clearly searching for them. These policies are enforced through a combination of human moderation and machine learning that detects problematic content for further review. For example, YouTube runs classifiers across videos looking for unusually high numbers of flesh coloured pixels. Such proactive detection technology is proving highly effective; in the third quarter of 2020, Facebook proactively removed 98.2% of adult nudity sexual activity content before it was flagged by users[6]. These policies are also reflected in members' advertising policies. Google Search does not generate revenue from, nor allow hyperlinks that drive traffic to, commercial pornography sites, nor does it allow pornography ads on search, or run Google ads against pornographic websites. On social media and content platforms, all members have advertising strict rules regarding pornography, adult products and services, and nudity.

In practice, the response times for the removal of publicly available RC and X18 content by major platforms once reported is extremely fast and well within the 24 hour proposal. That said, more than 24 hours may be required in certain cases that require more complex technical solutions, investigation, and for companies of varying sizes. As with all forms of content explored in the discussion paper, the Government might consider outlining a best practice timeframe for removal, an acceptable timeframe and clarify the legal position of intermediaries in cases that may necessitate more than 24 hours for content to be removed.

7.2 Removal notices to a hosting service provider

We note that the Bill extends the existing scheme under the EOSA to more services; we therefore reiterate the concerns and reiterate Recommendations 1-5 detailed in Section 2 relation to limiting the scope of services covered.

7.3 Revocation of removal notice

Section 113 indicates that the Commissioner may, by written notice given to the provider, revoke the removal notice. It is unclear why the revocation notice only pertains to removal notices under the online content scheme, and not to the other removal notices under the cyberbullying or image-based abuse scheme. There may be circumstances, such as after assessment of information provided to the Commissioner by a digital service provider, that may see the Commissioner wish to alter their assessment that led to the removal notice under those other schemes. Therefore, there must be provisions to similarly allow for the revocation of those removal notices.

Recommendation 22-23

22. The Commissioner's ability to issue a revocation notice should also be extended to all other takedown schemes under the Bill.

[6] *Facebook Community Standards Enforcement Report*, published February 2021, accessed at https://transparency.facebook.com/community-standards-enforcement#adult-nudity-and-sexual-activity

23. In relation to RC and X18+ content, the Government might consider outlining a best practice timeframe for removal, an acceptable timeframe and clarify the legal position of intermediaries in cases that may necessitate more than 24 hours for content to be removed.

8. Industry codes and standards

8.1 Industry codes

We welcome the emphasis on industry codes in the Bill. Issues relating to protecting and promoting online safety are highly complex, and benefit immensely from the extensive practical experience that relevant digital services bring from their own efforts in this area.

As mentioned, DIGI is a non-profit industry association that advocates for the interests of the digital industry in Australia, with Google, Facebook, Twitter and Verizon Media as its founding members. DIGI also has an associate membership program and our other members include Redbubble, eBay, Change.org and GoFundMe. 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. Our current membership includes key sections of the online industry, and we would welcome the opportunity to collaborate with the eSafety Commissioner on the development of such industry codes.

For an industry association, the development of an industry code comes with a significant investment of time, expense and adjustment of other priorities. We are concerned with the requirement in the Bill that different sections of the digital industry produce industry codes within six months of the passing of the law. Based on our direct experience, we suggest that this timeframe is unrealistic. Firstly, it may not allow enough time for the Office to develop clear written guidance to the industry about the scope of the code. It is critical that there is a clear understanding of the specific responsibilities, deliverables and lead times required by both regulators and industry participants if a set timetable is to be achieved with the development of an industry code. DIGI's experience is that a workable code requires a minimum of 12 months to develop, from the time that the Commissioner releases written guidance on the code for the industry. This will ensure the industry has sufficient time to engage in a thorough public consultation process.

By way of example, DIGI is preparing to launch an industry code of practice in response to the Government's request, announced in December 2019, that the digital industry to develop a code of practice on how digital products and services would address disinformation. We provide detail of the timeframe of our process to inform consideration of a more realistic code development process timeline.

Phase 1: Agreement, resourcing & partner engagement

- In March 2020, DIGI reached agreement within its membership that it would lead the development of this process.
- In April, DIGI contracted University of Technology Sydney's Centre for Media Transition (UTS CMT) and First Draft as key academic and civil society partners to support the development of the Code. UTS CMT is an interdisciplinary research centre that investigates key areas of media evolution and digital transition. First Draft is a global organisation that empowers societies with the knowledge, understanding and tools needed to outsmart false and misleading information.
- DIGI also convened a wider industry committee of potential signatories, outside of DIGI's current membership, to support the development of the code.

Phase 2: Issues mapping

- The Government tasked the ACMA with oversight of the code on December 11, 2019 and on June 26, 2020, the ACMA released a discussion paper outlining its expectations of the code.
- UTS and First Draft interviewed members of the industry committee, and conducted research into disinformation and misinformation in Australia. They also conducted a review of regulatory responses in different jurisdictions, and industry responses to the challenges.

- This research was released publicly as a discussion paper. This discussion paper commissioned by DIGI and the ACMA's discussion paper were used to inform the development of the draft code.

Phase 3: Initial draft development

- A draft code of conduct was developed, and refined with input from the industry committee for comment over the course of August and September.

Phase 4: Public consultation

- On October 19, a six week public consultation was launched ending on November 24, 2020.
- The code was made publicly available on the DIGI website, and was open for submissions from the general public.
- During this consultation, DIGI, UTS and First Draft proactively identified interested civil society, consumer and academic stakeholders. They were emailed the draft code and invited to comment. A smaller subset of this group was also invited to offer their views on the code at a roundtable meeting.

Phase 5: Revisions and final report

- All submissions were closely reviewed.
- Input from submissions was summarised into a report, later updated to indicate where the feedback had been reflected.
- Input was also sought from the ACMA.
- Over the course of December and January, the draft code was updated to reflect all stakeholder input.

Phase 6: Adoption

- In order to adopt a code, potential signatories must undertake an internal approval process, generally involving cross-functional institutional review, in order to determine whether they can become official signatories.
- At the completion of this process, which DIGI expects to be later in February 2021, the code will be launched with an initial list of signatories.

Phase 7: Ongoing administration

- After the initial development of the code, a system for the ongoing administration of the code must be developed and maintained.

As the overview of our process to develop the code of practice on disinformation illustrates, for an industry association to manage a robust and consultative code development process takes time. A six month process, beginning the date of the passage of the law, would not enable such a consultative and robust code development process. It also takes time for the regulator overseeing the code to issue guidance to industry on the code, which in this case took over six months. We therefore recommend a minimum of 12 months, from the date at which the Commissioner is able to release information on what is expected to be covered in the code.

The Bill in current form offers a long list of examples of matters that may be dealt with by industry codes and industry standards, in Section 138; this list is not sufficient guidance for an industry association to embark upon code development, nor to be set up for success in this endeavour. Furthermore, if this list of examples in Section 138 is intended to be an exhaustive list of expectations of code components, then this is far too ambitious for industry to achieve within six months. Given that the expectation seems to be that different sections of the online industry develop codes pertaining to their own services, different industry associations may need to coordinate these processes and move them forward in parallel; a coordination process that requires time.

Finally, we also note that the Bill affords the eSafety Office 12 months for the development of industry standards; it is unclear why the Office is afforded more time than industry associations for this comparable task.

As mentioned, we welcome the opportunity to develop industry codes in relation to online safety. However, we seriously urge the Department to ensure that the timeframe is realistic to ensure the robust and consultative process that is required to develop an effective regulatory response that relevant companies can adopt.

8.2. Industry standards

We understand that the Bill enables the Commissioner to, by legislative instrument, determine a standard that applies to participants in a particular section of the online industry.

The relationship between the industry standards and industry codes under the Bill is wholly unclear. The only connection identified is that industry standards prevail over industry codes, should there be inconsistencies between the two. It remains unclear whether an industry standard will be enacted upon an assessment that the industry code is not proving effective in addressing the challenges it seeks to address. It is also unclear what role the Office would play in the administration of the code, or in oversight of any changes that may need to be made to it. Clarification on these matters is needed for industry.

Furthermore, there is no requirement that the Commissioner evaluate any relevant industry code(s) prior to determining an industry standard. This means that an industry association, and digital service providers, would invest significant time and resources in developing an industry code, only to have it overruled at any time, without explanation. This creates uncertainty and confusion for a potential signatory, or any stakeholder, participating in good faith in a code development process.

Finally, the ability to enact industry standards simultaneously to the enactment of industry codes pertaining to the same issue, is inconsistent with other processes administered by other regulators. To use the example above of the disinformation code, the industry code was developed in response to the Government's agenda outlined in *Regulating in the Digital Age: Government Response and Implementation Roadmap for the Digital Platforms Inquiry*, which was developed following the ACCC's Digital Platforms Inquiry. The Roadmap states:

> *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 Government will assess the success of the codes and consider the need for any further reform in 2021.*

Under this process, industry associations and potential signatories clearly understand the timeframe and the milestones where the effectiveness of the code of practice will be reviewed.

Recommendations 24-26

24. Extend the deadline for the development of industry codes to at least 12 months from the date that the eSafety Office releases guidance on its expectations of the code.

25. The Bill must include a reasonable timeframe and evaluation process of an existing industry code before the Commissioner is able to enact an industry standard that overlaps or indeed overrules the content of the industry code.

26. There must be detail provided about how the Commissioner will strive for procedural fairness in both determining the need for an industry standard, as well as the process of developing the industry standard itself.

9. Role of the eSafety Commissioner

9.1. Assessment of needs and gaps

The eSafety Office plays an important role in giving Australians one point of contact and extremely valuable educational information about online safety, and is a key partner in DIGI members' efforts in this area. DIGI members have a productive working relationship with the current Commissioner, Julie Inman-Grant. We note also that there acknowledgement in the Discussion Paper that she has observed the prompt removal times online service providers have achieved on a voluntary basis.

9.2. Discretionary power of the Office

We are very concerned that the Bill's proposals vest a high amount of discretion within the eSafety Commissioner's Office. The Bill allows the eSafety Commissioner to issue removal notices for content that is not illegal, nor well defined or understood under Australian law. The Bill allows the Commissioner to demand transparency reports in relation to the BOSE without a mechanism for providers to challenge the reports for not meeting the specified criteria. The Bill allows the Commissioner to overrule industry codes with industry standards at any time, without explanation.

The Bill even goes as far as to allow the Federal Court to order a person to cease providing a social media service, electronic service, designated internet service or internet carriage service based on an assessment from the Commissioner that there were "there were 2 or more occasions during the previous 12 months on which the person contravened a civil penalty provision of this Part."

We welcome the fact that the Bill allows for application to be made to the Administrative Appeals Tribunal for a review of a decision of the Commissioner. However, providers and the public cannot rely on an appeals process alone to ensure that the Commissioner's powers are exercised consistently and fairly, particularly when the current practice of the Office is to resolve issues through informal industry communication, without exercising legislative powers. While we believe the Office's functions are important, and it must be empowered to act quickly in the area of online safety, there needs to be consistency and procedural fairness in the decision-making behind removal notices, service provider notifications, transparency report demands, industry standards and applications for cessation of services to the Federal Court.

9.3 Inconsistency with other regulators

We also note that the powers afforded to the eSafety Office are inconsistent with other regulators. The Bill states that "for the purposes of the finance law (within the meaning of the Public Governance, Performance and Accountability Act 2013), the Commissioner is an official of the ACMA". It is therefore fitting that the eSafety Office's accountability framework be strengthened to provide for a measure of independent oversight of the exercise of its regulatory powers. This should include the extent to which it has appropriate governance arrangements in place, requiring publicly accessible documentation about its decision-making processes, including the extent to which decisions required the exercise of informal powers under legislation. Consideration could be given to adopting a committee or "authority" tasked with this function similar to those set up within the ACMA.

These accountability models are vitally important in ensuring public trust in the regulator's decisions. While one may expect the current Commissioner may reasonably and sensibly exercise their powers today, the Bill needs to allow for changes in personnel at the Office in the future and mitigate the possibility that these powers could be abused.

9.4. Transparency reporting

It is important that digital providers and the public have transparency into the decision-making processes of the Office. We would recommend some form of transparency reporting into the notices served, both to platforms and to end users. This will assist efforts to evaluate the effectiveness of the

current and proposed schemes under the Act, any gaps in those schemes, and whether there is a need for increased powers for the Commissioner as a result.

Recommendations 27-30

27. While we are supportive of modernised laws that keep pace with the challenges of online safety, we encourage exploration of the specific needs that are not currently being met under the current scheme -- such as content categories, or sections of the industry where greater collaboration is necessary -- along with contemplation of more targeted provisions in the Bill to address these defined problems.

28. There must be documented systems and processes of procedural fairness in the decision-making behind removal notices, service provider notifications, transparency report demands, industry standards and applications for cessation of services to the Federal Court.

29. The Government should consider accountability models that exist within other regulators, such as the ACMA, to ensure oversight of the eSafety Office's discretionary powers and decision-making practices.

30. A form of regular transparency reporting by the eSafety Office's on their decision-making practices, and the notices served, should be introduced. This would provide public accountability, and a strong basis from which evaluate the success of the reform program.