Friday February 21, 2020

Director, Online Safety Research and Reform Section
Department of Infrastructure, Transport, Regional Development and Communications
By email: onlinesafety@communications.gov.au

Dear Director,

Thank you for the opportunity to engage with the Australian Government about its proposed Online Safety Act as outlined in the *Online Safety Legislative Reform* discussion paper released in December 2019.

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 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 with 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. This submission articulates some of the implementation questions that may arise for the digital industry in the proposals outlined in the discussion paper, which we hope can be taken into account as the Act is developed.

DIGI looks forward to further engaging with the Online Safety Act reform process. Should you have any questions or 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

Objects of the new Act

Discussion paper questions:
1. Are the proposed high level objects appropriate? Are there any additions or alternatives that are warranted?
2. Is the proposed statement of regulatory policy sufficiently broad to address online harms in Australia? Are there aspects of the proposed principles that should be modified or omitted, or are there other principles that should be considered?

As noted, 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 also supportive of efforts to streamline legislation

pertaining to online safety under one consolidated Online Safety Act, and that such an Act include an objects section to aid in the understanding of its purpose and from which industry and other stakeholders can assess its effectiveness.

Preventing online harms

DIGI believes high level object of "preventing online harms" is not necessarily 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. Should the Government decide to include "preventing online harms" in the final version of objects of the Act, we encourage the consideration of more behavioural and perpetrator level policy approaches that are better suited to this goal. It is worth noting that the current *Enhancing Online Safety Act* (EOSA) enables the Office of the eSafety Commissioner (eSafety 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.

Furthermore, related to the proposed high level object of "preventing online harms", the Australian Government made an election commitment on May 5 2019 to increase maximum penalties for using a carriage service to menace, harass or cause offence[1]. It 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 are important ways to deter child sexual exploitation and cyberbullying, consistent with the proposed object, yet it is surprising that they do not feature in the discussion paper.

Hate speech

In relation to the breadth of online harms, DIGI encourages further attention on the issue of hate speech. While DIGI members all have set policies to restrict hate speech at a global level, Australia has a definition of hate speech under the *Racial Discrimination Act 1975* (Cth) that only applies to race-based hate speech, and does not include religious-based or gender-based speech. Frameworks to ensure online safety provide an opportunity to establish better legal foundations to combat hate speech; we therefore encourage the development of a clearer legislative framework that defines hate speech to assist enforcement agencies and prosecutors with how to define and approach this issue. 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.

Private messaging, email & enterprise software

Across all of the recommendations in the proposed Act, further clarity is needed on the scope of digital services that may be covered under "designated Internet services", as well as "hosting services" and "relevant electronic services". The final Act should define these clearly with examples, and consider the varying capability of different types of services to address the identified content in a targeted fashion. Email, private messaging and enterprise software are services where end users have a greater expectation of privacy and also

[1]Liberal Party of Australia (5/5/2019), Keeping Australians Safe Online, accessed at https://www.liberal.org.au/latest-news/2019/05/05/keeping-australians-safe-online

greater control through tools, such as communication blocking or administrator intervention. Therefore, we encourage a reconsideration of whether such private messaging services are subject to the proposals to the same extent as publicly available posts or other content. We note that the *General Scheme of the Online Safety & Media Regulation Bill 2019* currently being examined in Ireland differentiates between publicly available online content and private messaging, per the excerpt below:

> *However, in relation to two examined categories, private communications services and private online storage services, it is provided that the Media Commission's code making powers in relation to these services be explicitly limited to matters relating to content which it is a criminal offence to disseminate. The reason for this is that these services raise particular rights balancing issues, especially regarding the right to privacy, which make it difficult to justify giving the Commission to power to require them to take measures in relation to non-criminal harmful online content*[2]*.*

Basic online safety expectations (BOSE)

Discussion paper questions:
3. Is there merit in the BOSE concept?
4. Are there matters (other than those canvassed in the Charter) that should be considered for the BOSE? Are there any matters in the Charter that should not be part of the BOSE?
5. What factors should be considered by the eSafety Commissioner in determining particular entities that are required to adhere to transparency reporting requirements (e.g. size, number of Australian users, history of upheld complaints)?
6. Should there be sanctions for companies that fail to meet the BOSE, beyond the proposed reporting and publication arrangements?

Avoiding duplication

DIGI's members share the Government's goal and expectation that technology companies and digital platforms should be proactive in embedding online safety at the outset.

We have welcomed the Safety by Design initiative by the eSafety Office, both to mitigate and to address the wide range of safety challenges that occur through digital products and services. The Safety by Design principles reflect the online safety work of our member companies and also provide a roadmap for companies working to be responsible players in this space. DIGI has welcomed the opportunity of working collaboratively with the eSafety Office in consulting on these principles, and we look forward to continuing to consult on phase two of the Office's process.

The Basic Online Safety Expectations (BOSE) appear highly duplicative with the Safety by Design principles, as well as the Online Safety Charter released in December 2019, and risk creating confusion for the industry as to which principles they must follow. We believe there is merit in outlining a best practice expectation and roadmap for investing in user safety and

[2] Department of Communications, Climate Action and Environment, Ireland, *General Scheme of the Online Safety and Media Regulation Bill* accessed at https://www.dccae.gov.ie/en-ie/communications/legislation/Documents/154/General_Scheme_Online_Safety_Media_Regulation_Bill.pdf, p.97

associated transparency, that draws on learnings from the industry, but articulations of best practice should not form the basis of regulation. The Safety by Design principles and/or the Online Safety Charter, or consolidation of the two, might be the best means to communicate these best practice expectations outside of any legislation. We would encourage the Government to consider how these expectations are well socialised and understood widely across the broader digital industry.

Expectations are not "basic"

Having noted the concerns around duplication, we note that comments are sought through this consultation process on whether the Government's Online Safety Charter is a sufficient basis for the BOSE, or if other additional matters should be addressed. While the Charter is largely reflective of the work of large technology companies and DIGI's founding members, the expectations outlined should not be characterised as "basic". In particular, the features of the Charter outlined below relating to proactive content detection technology, external expert consultation and transparency reporting are not always accessible to the broad range of companies that host user-generated content:

> *Put processes in place to detect, surface, flag and remove illegal and harmful conduct, contact and content with the aim of preventing harms before they occur…. Where feasible and appropriate to the service, utilise technology to 'fingerprint' content that has been identified as illegal or harmful and deploy systems to prevent the attempted upload, re-upload or sharing of this material.*
>
> *...Carry out open engagement with a wide user-base, including experts and key stakeholders on the development, interpretation and application of safety standards and their effectiveness or appropriateness.*
>
> *...Publish an annual assessment of reported abuses on the service, alongside the open publication of meaningful analysis of metrics such as abuse data and reports, the effectiveness of moderation efforts and the extent to which community standards and terms of service are being satisfied through enforcement metrics.*[3]

These initiatives, while reflective of practices in a small number of industry-leading technology companies, should not be codified into binding standards under the Act for a range of reasons. In relation to content detection technology, well-intentioned and meaningful efforts to mitigate the upload of harmful content can fail despite a large investment in such technology because they have a margin for error, particularly when deployed in real time and when encountering emerging harmful content variations that may not have been seen before. Secondly, there are few "plug and play" technologies, such as Google's Jigsaw, that are commercially available solutions to different digital products and services and, where they exist, they function for some content types but not others. Thirdly, the development, maintenance and ongoing optimisation of product-customised technology is cost-prohibitive and arguably inappropriate for an industry-wide binding standard. It is also worth taking into account that smaller platforms might deal with content moderation differently to the larger platforms that rely on such technology, where a greater focus on human review may be appropriate given the smaller volumes of content; it may not be worth the significant investment in such technology if it is not suited to the size or structure of their platforms.

---

[3] Department of Infrastructure, Transport, Regional Development and Communications (2019), *Online Safety Charter*, accessed at https://www.communications.gov.au/documents/online-safety-charter-0

Transparency reporting

While large technology companies, including DIGI's founding members, release regular transparency reports, a global analysis of transparency reporting by AccessNow found that only 70 telecommunications and Internet companies have released transparency reports worldwide, and Telstra is the only Australian originating company to do so[4]. This analysis indicates that transparency reporting is a significant undertaking, particularly for small and medium sized businesses, and that more incremental approaches to encourage and support companies with transparency reporting may be needed in the first instance. We would speculate that the lack of transparency reporting is due to the significant amount of data collection, analysis and operational work that is required in order to produce such a report on an annual or more frequent basis; therefore it may better to encourage further conversations with industry in advance of regulation being considered.

Revenue, size, history, technical difficulty and nature of services provided are important considerations to consider when determining thresholds for transparency reporting; yet, as currently drafted, the proposal suggests that the eSafety Commissioner would have the discretion to determine particular entities that will be subject to the requirements. A better approach would be to encourage reporting against certain principles-based criteria that can apply across a set of diverse digital products and services. Furthermore, while efforts to ensure interoperability with overseas schemes such as the OECD transparency reporting protocol and UK Online Harms Paper are welcome, particularly for global platforms, it is worth noting that neither of these processes are currently complete and it may be premature to emulate specific proposals from within these processes when they remain unresolved and are being actively debated. Based on the outcome of these processes, the Government might consider a form of "mutual recognition" so that global or other country reporting is recognised as compliant for Australian purposes, noting that many digital products and services are global in nature.

Sanctions

We note that "the Government is not proposing to impose sanctions for non-compliance with the proposed basic online safety expectations at this stage, though reserves the right to explore this option in future if expectations are not being met." It is unclear what would constitute "expectations not being met". It would be unreasonable to see a situation where, if one company is a repeat offender, the whole industry then faces the prospect of sanctions as a result, and we caution the Government against such an approach. This posturing also creates regulatory uncertainty for businesses.

Cyberbullying scheme for children

Discussion paper questions:
7. Is the proposed expansion of the cyberbullying scheme for children to designated internet services and hosting services, in addition to relevant electronic service and social media services, appropriate?

[4] AccessNow (2019), *Transparency Reporting Index*, accessed at https://www.accessnow.org/transparency-reporting-index/

8. Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?
9. What are the likely compliance burdens of the proposed changes to the cyberbullying scheme on small and large businesses?
10. What other tools could the eSafety Commissioner utilise to effectively address cyberbullying in the circumstances where social media service and end-user notices are not well suited to the particular service upon which the cyberbullying has occurred?

DIGI shares the Government's 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.

Recognising that the cyberbullying of children can take place and must be addressed on a wide range of digital services -- not just those currently participating in the tiered scheme -- we believe that the expansion of the scheme is appropriate however it requires further clarity on the scope particularly in relation to "designated Internet services". We would appreciate a deeper understanding of the types of services that the Government envisages will fall under this category, how services will be designated and by whom, particularly in comparison with the new category of "ancillary service providers" discussed later in this submission. Is an Internet search engine, for instance, a designated Internet service or an ancillary service provider, or both? Is a web browser a designated Internet service? While our members promptly respond to notices served under the EOSA law, and have rapid response protocols in place with the eSafety Office, we do see merit in clarifying intermediary liability in exceptional cases where investigation of a complaint may take 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 outlined in the following section in relation to the cyberbullying scheme for adults where there are more frequently factors that may necessitate a longer timeframe.

Cyberbullying scheme for adults

Discussion paper questions:
11. Is the proposed application of the cyberbullying and cyber abuse schemes to designated internet services and hosting services, relevant electronic service and social media services, appropriate?
12. Is the proposed take-down period of 24 hours reasonable, or should this require take-down in a shorter period of time?
13. Do the proposed elements of a definition of adult cyber abuse appropriately balance the protection from harms with the expectation that adults should be able to express views freely, including robust differences of opinion?
14. Should the penalties differ under a cyber abuse scheme for adults and the cyberbullying scheme for children?

15. What additional tools or processes, in addition to removal notices, could be made available to the eSafety Commissioner to address cyber abuse occurring across the full range of services used by Australians?

Private messaging, email & enterprise software

As noted, further clarity is required on the scope of the proposed recommendation, particularly in relation to “designated Internet services”, as well as “hosting services” and “relevant electronic services”. While we appreciate the cyberbullying of adults can often have a private dimension, services such as email, private messaging and enterprise software are ones where end users have a greater expectation of privacy and also have greater control through tools, such as communication blocking or administrator intervention. Therefore, we encourage a reconsideration of whether such private messaging services are subject to the proposals to the same extent as publicly available posts or other content in line with other global approaches to such as that which is being considered in Ireland.

Illegal vs. other content

In relation to the proposed time frames of responding to notices, the discussion paper states that the shortening of take-down expectation times to 24 hours 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. Content that is not clearly illegal under NetzDG is subject to a seven day review period. There is no discretion for any regulatory body in Germany to deem content “harmful” under NetzDG and demand that such content is removed within 24 hours on that basis.

In this regard, we would argue the eSafety Commissioner should operate to uphold Australian law; it is not appropriate that the eSafety 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. 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. As noted above, the Australian Government might wish to examine areas of harmful content that are commonly covered under digital platforms’ Terms of Service, such as hate speech, that expose gaps within Australian law and consider opportunities for law reform.

Relatedly, the proposal that the Commissioner can request account restrictions may pose implementation challenges at the service level; for example, some digital platforms offer a “single sign on” across multiple products, thereby making an account restriction on one service only impossible to implement. That said, the service may make the determination itself that a particular user has violated its Terms of Service to the extent that account restrictions become necessary -- for example, some platforms may have a “three strikes” policy whereby three violations of the Terms of Service trigger account restriction. That is again to say, the determination around account restriction should be made in relation to Terms of Service, while the regulator’s role should be focused on ensuring the prompt removal of content that is illegal.

The regulator 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.

Statutory test for adult cyberbullying

The element of the statutory test based on the conclusions of “an ordinary reasonable person” is extremely challenging to interpret. In practice, digital platforms have far more granular considerations when assessing the cyberbullying of adults. For example, some of their considerations might include questions like: Does the content attack someone for their private positions vs. public opinions or actions? Does the content attack 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. Given that the proposal indicates that the statement of regulatory policy would indicate that the Act is seeking to “balance the competing objectives of user safety and freedom of expression”, we suggest that this definition of the cyberbullying of adults be reconsidered. Unlike the cyberbullying of children, where responsible digital platforms 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.

24 hour timeframe

It is important to emphasise that NetzDG does not always require a 24 hour takedown of content by platforms subject to the law. The time frame 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. 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 code has been violated, the content has to be assessed more carefully. 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.

The reality is that most clear, prima facie examples of cyberbullying will be removed well within 24 hours on DIGI member platforms. The discussion paper even acknowledges that the eSafety Commissioner has observed the prompt removal times online service providers have achieved on a voluntary basis. However, noting some of the complexities explained above, 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 24 hours in such cases. We welcome further information from the Government justifying the need to move to a 24 hour turnaround time for content removal and the rationale for this choice of metric. Imposing a short and prescriptive turn around time for content removal will be particularly impactful on smaller digital platforms, upon which this proposed Act may have a disproportionate burden, as they do not have the same technical and human resource capability for trust and safety operations as larger digital platforms. Analysis of the

transparency reports from NetzDG indicates that the expense of implementing NetzDG was high for the smaller platforms to which it applied[5].

Defamation overlap & law reform

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 "Stage 2" process occurring in late 2020 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 (e.g. A statement where one person calls another a fascist, can be considered bullying and defamation). Definitions need to be clearly delineated to provide meaningful guidance to platforms about their responsibilities under the adult cyberbullying scheme and the outcome of the Stage 2 defamation law reform process in relation to digital platforms.

Consistent with emerging 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 way -- 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 turn around time as the measurement of compliance under any new Online Safety Act, it ought to also provide legal protection for organisations where there are legitimate circumstances that mean that reviewing and responding to the complaint may take longer.

Penalties

As previously noted, we encourage the Australian Government to consider the role of end-user notices in actually preventing cyberbullying of children and all other forms of abuse outlined in the proposed Act. We therefore welcome the proposal that the establishment of a cyber abuse scheme for Australian adults would include an equivalent end user take-down and penalty regime. That said, the penalty regime should recognise that the nature of relationships between adults can be much more complex than between children. Therefore a significant amount more work is needed to design a threshold that recognises these complexities, such as the aforementioned considerations around public figures and debates that are arguably in the public interest.

Image-based abuse scheme

Discussion paper questions:
16. Is the proposed take-down period for the image-based abuse scheme of 24 hours reasonable, or should this require take-down in a shorter period of time?
17. Does the image-based abuse scheme require any other modifications or updates to remain fit for purpose?

[5] Jason Pielemeier of the Global Network Initiative, (27/2/2019), *NetzDG: A Key Test for the Regulation of Tech Companies*, accessed at https://medium.com/design-and-tech-co/netzdg-a-key-test-for-the-regulation-of-tech-companies-e4ba205b566c

18. What additional tools or processes, in addition to removal notices, could be made available to the eSafety Commissioner to address image-based abuse being perpetrated across the range of services used by Australians?

In practice, 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 above, 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. As with all forms of cyber abuse outlined in the discussion paper, the Government might consider outlining a best practice timeframe, an acceptable timeframe and clarify the legal position of intermediaries in cases that may necessitate more than 24 hours for content to be removed.

As noted, we encourage the consideration of end-user notices and penalties and the promotion of these to raise awareness of the criminal nature of image-based abuse, and to deter the occurrence of it.

Illegal & harmful content scheme

Discussion paper questions:
19. Is the proposed application of the take-down powers in the revised online content scheme appropriate?
20. Are there other methods to manage access to harmful online content that should be considered in the new Online Safety Act?
21. Are there services that should be covered by the new online content scheme other than social media services, relevant electronic services and designated internet services?
22. Is the proposed take-down period of 24 hours for the online content scheme reasonable or should this require take-down in a shorter period of time?

RC and X18+ content violates most responsible digital platforms' Terms of Service; all DIGI members have strict content policies in relation to pornographic content and child sexual exploitation material. 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. 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; last quarter, Facebook proactively removed 98.4% of adult nudity sexual activity content, and 99.5% of child nudity and sexual exploitation content, before it was flagged by users.

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 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. Some platforms have also introduced technology that detects nudity and other indicators to pre-emptively remove such content without reliance on user reporting. 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.

Parental controls & accreditation scheme

Discussion paper questions:
24. To what extent would an expanded accreditation scheme for opt-in tools and services assist parents and carers in mitigating the risk of access by minors to potentially harmful content?
25. What categories of tools and services should be included in an accreditation program, aside from content filters?
26. What are the likely costs of developing and maintaining an accreditation scheme for opt-in tools and services to assist parents and carers in managing access to online content by minors?

Opt-in tools and services are often most effective when tailored to a specific platform, and today the digital industry offers a range of parental controls to offer minors a safer and more controlled experience online. DIGI members have a range of tools to protect the experience of minors online. Google's Safe Search filter prevents ads containing or promoting nudity, sexually suggestive content, adult entertainment and other services from appearing within search results. Google's Family Link app is pre-installed on all new Android devices and enables parents to monitor, approve and restrict access to certain apps and websites on a child's device. Twitter also has Quality Filter and Sensitive Media settings for sensitive content, which places images and videos behind an interstitial warning message, that needs to be acknowledged before flagged media on Twitter can be viewed; using this feature means that people who don't want to see sensitive media can avoid it, or make an informed decision before they choose to view it. Facebook uses strict privacy and visibility default settings for people between the ages of 13 and 17. Given the existence of these activities, in addition to the service provider level tools explored in the discussion paper under Table 2, it may not be worth the significant Government investment to operate an accreditation scheme when there is insufficient data that demonstrates that such schemes actually inform meaningful parental choices.

Furthermore, the proposal for the "best available technology solutions" is unclear. As noted, should the implication here be that this involve content filters, this raises several cost and accuracy challenges with content detection technologies which have been previously noted in the section on the BOSE. Additionally, as noted in the Classification Review discussion paper, the definition of 'film' in the Classification Act is broad and technically covers all content online apart from online games and online advertisements. We agree with the assessment in that discussion paper that "the definition of 'film' be clarified so that industry

has clearer obligations about what must be classified, as it is impractical that virtually all online content must be classified in the same manner."[6].

If the requirement to use the "best available technology" is retained, the costs of the accreditation scheme across the digital industry need to be considered so as to avoid being prohibitive for small to medium sized businesses. The costs of this scheme also need to be examined alongside the costs of other proposed possible industry-funded schemes, such as the Government's proposed pilot of a dispute resolution scheme in response to the ACCC Digital Platforms Inquiry.

Ancillary service provider notice scheme

Discussion paper questions:
31. Is there merit in the concept of an ancillary service provider notice scheme?

There is merit to the concept as long as it is a notice and takedown scheme without proactive monitoring requirements, and includes requirements for the eSafety Office to approach the website host in the first instance, prior to approaching the ancillary service provider with a notice. The proposal as outlined requires some further clarification on these points. It is also worth noting that some search engines, such as Google, have existing processes whereby regulators can request a link be removed from because it violates Australian law.

Role of the eSafety Commissioner

Discussion paper questions:
36. Are the eSafety Commissioner's functions still fit for purpose? Is anything missing?
37. To what extent should the existing functions of the eSafety Commissioner be streamlined? Are there particular functions that need to be maintained, or new functions that should be specified?
38. To what extent should the functions of the eSafety Commissioner be prioritised?
39. What are the likely impacts, including resource implications, on other agencies and businesses of a new Online Safety Act?

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.

While we are supportive of modernised laws that keep pace with the challenges of online safety, it is worth noting that the proposals under the proposed Online Safety Act vest a high amount of discretion within the eSafety Commissioner's Office, which is inconsistent with other Australian regulators. As noted, the proposals that the eSafety Commissioner can issue takedown notices for content that is not illegal under Australian law, can make determinations about a company's own discretionary Terms of Service and prescriptions

[6] Department of Infrastructure, Transport, Regional Development and Communications (2019), *Review of Australian classification regulation—discussion paper*, accessed at https://www.communications.gov.au/have-your-say/review-australian-classification-regulation

about the technology that it may use will all involve a significant amount of discretion being exercised by the Commissioner and the Office's staff; discretion that we may trust the Commissioner to reasonably and sensibly exercise today but does not account for changes in personnel at the Office in the future.

While the eSafety Office's functions are important and it must be empowered to act quickly in the area of online safety, there must also be procedural fairness for the notices to be reviewed and appealed. There is a need for a clear process to review any errors made in the issuing of notices. For example, the final Act might consider allowing intermediaries to make an application to the Administrative Appeals Tribunal for the review of any decisions made by the eSafety Commissioner to give a content service provider a notice. This is an important check and balance that is consistent with the statement of regulatory policy for the proposed Act that it "balance the competing objectives of user safety and freedom of expression". In general, increased oversight and accountability of the Office is appropriate given the proposed increase in powers.

In closing, we welcome the acknowledgement in the discussion paper that the eSafety Commissioner has observed the prompt removal times online service providers have achieved on a voluntary basis. While we are supportive of efforts to consolidate reform in online safety into a single Act, we encourage further 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 more targeted solutions to address these defined problems.