COMMUNICATIONS
ALLIANCE LTD

Communications Alliance

Submission

to the

Department of Communications and the Arts

(Department of Infrastructure, Transport, Regional Development and Communications)

Online Safety Legislation Reform – Discussion Paper

19 February 2020

CONTENTS

1. ABOUT COMMUNICATIONS ALLIANCE

Communications Alliance is the primary telecommunications industry body in Australia. Its membership is drawn from a wide cross-section of the communications industry, including carriers, carriage and internet service providers, content providers, equipment vendors, IT companies, consultants and business groups.

Its vision is to provide a unified voice for the telecommunications industry and to lead it into the next generation of converging networks, technologies and services. The prime mission of Communications Alliance is to promote the growth of the Australian communications industry and the protection of consumer interests by fostering the highest standards of business ethics and behaviour through industry self-governance. For more details about Communications Alliance, see http://www.commsalliance.com.au.

In March 2014, Communications Alliance assumed responsibility for the industry codes and core responsibilities of the Internet Industry Association (IIA) (which was in the process of dissolving). Consequently, Communications Alliance became the owner of the IIA industry codes, including the *Hosting Content Within Australia Code*, the *Providing Access To Content Hosted Within Australia Code* (together the Internet and Mobile Content Codes) and the *Content Services Code*. Communications Alliance also took over responsibility for the Family Friendly Internet Filter scheme (FFF) scheme (including the Ladybird Logo).

2. INTRODUCTION

Communications Alliance welcomes the opportunity to make a submission to the Department of Communications and the Arts (now the Department of Infrastructure, Transport, Regional Development and Communications, Department) Discussion Paper *Online Safety Legislation Reform* (Discussion Paper).

The lives of Australians, and of most nations worldwide, are increasingly influenced by an online environment in which they actively or passively participate. Access to the internet is almost universal in most developed countries, and is increasingly considered to be a basic human right. The internet has not only become an essential tool for formal and informal education in all areas of society but is also a key mechanism for communication, engagement and even play. With one of the highest smart-phone penetration rates in the world and fast and reliable mobile internet in most of the populated areas of Australia, this online environment is now almost always available at our fingertips.

Australian governments have created rules, guidelines and behavioural expectations on how to keep individuals safe in our physical environments (e.g. on our roads, in maritime situations, national parks, etc.) while simultaneously enjoying these physical environments and ensuring that the ecosystem of this environment can remain intact and flourish. Much in the same way, our society must create and apply certain standards for our online environment to ensure the safety of its citizens and providing the conditions in which the online world can continue to evolve and furnish the services that we have come to love and depend on. A safe online environment is a shared responsibility of Government, Industry and users.

The communications industry recognises that access to some online content, particularly by minors or vulnerable adults, may have detrimental effects on the physical, social and emotional well-being of the user and may also influence their values with regards to sexuality, relationships, violence, security, racial and religious equality, tolerance and many other key societal values. The proliferation of online social networking poses additional challenges around cyberbullying and the unwanted sharing of (sometimes intimate) images. In this context, it must be noted that social media platforms and search engines dedicate vast amounts of time and resources to minimise abuse of their services and potential harm that may result from content that is accessible through their services. The overwhelming majority of abuses are detected and removed by the major platforms proactively and without requiring or using an internal or external escalation mechanism.

It goes without saying that illegal content, especially material relating to child sexual abuse and terrorism, must be eradicated to the extent possible and as quickly as possible, to minimise the detrimental effects on all parties involved.

As described in the Discussion Paper, Australia has existing mechanisms and tools in place within both Government and Industry, including the Office of the eSafety Commissioner, to address many of the issues described above.

We agree with the general premise that a review of some of the underlying legislative framework is timely – if not overdue – to ensure that the online world is governed by technology and platform neutral, practical and principles-based rules that, to the largest extent possible, are able to flexibly adjust to the dynamic environment that they pertain to.

As in the past, our industry continues to engage closely with all stakeholders, including enforcement agencies, and is keen to assist, where possible, to create, maintain and promote a safe online environment.

3. COMMENTS AND OBSERVATIONS

3.1. Co-regulatory approach

As already highlighted in our submission to the Reviews of the *Enhancing Online safety Act 2015* and the Online Content Scheme (OCS) in July 2018, we welcome the proposal to consolidate the various pieces of legislation and regulation that currently form the online safety framework, including Schedules 5 and 7 of the *Broadcasting Services Act 1992* (BSA), into a single piece of legislation.

We also agree with the premise that parts of the existing framework require updating. This is not only required to adequately reflect technological and societal developments over the past two decades but also to ensure that Australia's legislative and co-regulatory framework is sufficiently flexible to promptly accommodate any future change that will undoubtedly – and most likely at an even faster pace – occur in this area. The deployment of 5G, the burgeoning influence of the Internet of Things, progress in relation to virtual and augmented reality and the creation and widespread use of artificial intelligence may serve as examples of significant technological change already at our doorstep.

Consequently, we urge the Department to adopt a co-regulatory and principles-based approach to the largest extent possible to allow Industry to make the required adjustments to regulation as required by technological and societal developments. We equally caution against undue prescription in the proposed *Online Safety Act* (OSA), noting that the current prescriptiveness of Schedules 5 and 7 of the BSA have prevented a meaningful overhaul of the OCS for several years.

3.2. Overseas online content debate

The Discussion Paper appears to suggest the proposals put forward align with international legislation and thinking.

We believe this is only partly an accurate reflection. It should be understood that the UK is still hotly debating the approach put forward by the UK Government in the *Online Harms White Paper*, released in April 2019.While the UK Government is still consulting on potential measures resulting from the White Paper, it is clear that there remains significant work to be undertaken on these measures. In its recent response, the UK Government has already signalled a different approach than originally put forward in the White Paper in some areas, and there may be further adjustments to the proposals before the UK Government reaches a final position.

We equally highlight that the German *Netzwerkdurchsetzungsgesetz* (NetzDG) not only defines the content to be removed by reference to statute (i.e. the German Penal Code) but also involves longer timeframes for removal of content that is not manifestly illegal than indicated in the Discussion Paper. The application of the German legislation is also limited to a narrower set of platforms and services. For example, it excludes private messaging services. The law may also be subject to constitutional challenge in the future. Transparency reporting by the companies impacted by this law also demonstrates an extraordinarily high rate of requests to remove content that does not fall within the scope of the legislation (i.e. content that falls outside the German Penal Code). Consequently, the effectiveness of this law in achieving the public policy objectives can be questioned.

Overall it should be noted that most overseas approaches centre on the removal of hate speech, often defined in their respective legislations, and the worst illegal content rather than harmful content.

3.3. Submission, process and concurrent reviews

While a number of the proposals put forward in the Discussion Paper appear to be reasonable and are welcome, we raise concern with other aspects of the Discussion Paper, most notably any explicit or implied notion that would amount to online service providers, or any other stakeholder, excessively censoring (through removal or request to remove) legal content on the basis that it may be considered harmful for end-users. We are also concerned by the sometimes broad application of the proposals. These concerns are exacerbated by the proposed shortened reaction timeframes (24 hours).

In our submission, we will follow the structure of the Discussion Paper to provide our feedback on the proposed reforms.

We also note that, in order to ensure the actual legislation is practical, appropriate, balanced and fit-for-purpose, it is imperative that the consultation process allocates sufficient time for all stakeholders to comment on and cooperatively discuss the exposure draft of the new OSA and any associated documents.

With respect to concurrent reviews, we urge Government to allow sufficient time to appropriately consider the findings and recommendations of the review of the Model Defamation Provisions as it is likely that there will be significant overlap in content that may be deemed defamatory and fall under the adult cyber abuse scheme.

We acknowledge the Review of Australian classification regulation. At the time of writing, we do not anticipate providing substantial feedback to this review. However, we lend our support to the proposal to amend the definition of 'film' to exclude certain content, e.g. YouTube videos.

3.4. Basic online safety expectations (BOSE)

The Discussion Paper indicates that the Safety by Design (SbD) principles, the Online Safety Charter (Charter) released by Government in December 2019, and the outcomes of the consultation on the online safety legislation reform (Reform) will inform basic online safety expectations (BOSE or Expectations) for inclusion in the OSA.[1] It is therefore useful to consider the SbD and the Charter in this context.

We commend the eSafety Commissioner for the large amount of work that has gone into the development of the SbD principles, and the intensive consultation that has guided the development of those principles. Indeed, some very prominent platform/search engine providers have adopted the principles for the development and design of new services and technology.

As indicated in the Discussion Paper, it will be important to continue to work with Industry to not only spread the adoption of these principles but also to ensure the principles remain fit-for-purpose in a fast-paced environment and, importantly, to provide practical guidance on the implementation of such principles. The latter will be particularly important for smaller online service providers with limited experience and/or resources dedicated to safety considerations.

We also lend our in-principle support to the Online Safety Charter. However, the elements of the Charter that refer to an undefined concept of 'harm', e.g. "detect, surface, flag and remove illegal and harmful conduct, contact or content"[2] and the expectation that such 'harms' be "prevent[ed]… before they occur"[3] raise concern as they lack specificity and are likely to give rise to subjective or even arbitrary decisions on what is considered harmful. From

[1] p.12, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019
[2] p.6, *Online safety charter, A statement of the Australian Government's expectation of online service providers*, Australian Government, Dec. 2019
[3] p.6, *Online safety charter, A statement of the Australian Government's expectation of online service providers*, Australian Government, Dec. 2019

a rule of law perspective, Industry requires certainty on what they must comply with and any associated oversight mechanisms.

Illegal content, while still difficult to detect and often also subject to legal analysis where content is not manifestly illegal, is, as the term suggests, defined through law, and the application of this law is subject to review by the Courts. The concept of harm, however, appears to be subjective and variable on both sides of the equation: online service provides may have various and subjective notions of harm, in the same way as end-users will have widely varying understandings of what they feel is harmful. The matter is being complicated by the expectation that harms be prevented before they occur, meaning that the potential subject of the harm would not be available as a source of information as to whether harm has actually occurred. The Charter does not provide any guidance on the concept of harm, nor does it provide any indication of harm thresholds.

While this vagueness and lack of specificity may still be acceptable (and intentional) for a statement of expectation, compliance with which is voluntary (though note the 'big stick' if expectations are not being met!), it is concerning that the Charter will inform the BOSE which are to be included into enforceable law, i.e. the OSA. Later sections of the Discussion Paper make clear that it is envisaged that the concept of 'harmful content' not only plays a role in the BOSE but finds direct application in Codes that Industry is to develop under the proposed OSA. We will discuss these issues in Section 3.7 (*Addressing illegal and harmful content*) of our submission.

With regard to the BOSE themselves, we note a lack of specificity around the applicability of the Expectations. The Discussion Paper envisages the BOSE to apply to all social media services without further explanation as to the definitional aspects of such a service. For example, would there be a size (e.g. by subscriber numbers) threshold and/or would certain services such as services that are designed to facilitate communications between individuals (email, messenger services etc.) or services that for the purpose of sharing professional content be included in the definition of 'social media services'? The answer to those and other questions will fundamentally influence the discussion around the practical application of the Expectations and impact on the risk of potential unintended consequences of the BOSE.

Large social media service providers and search engines already voluntarily provide extensive reports that detail their online safety measures and provide metrics on content removal, complaints etc., i.e. transparency reporting already forms part of standard business practice for those organisations.

We welcome the intention to create a single reporting framework to minimise costs and regulatory overhead. However, any consolidation of existing reporting mechanisms and the establishment of new mechanisms must leave sufficient flexibility for service providers captured under those obligations to use metrics and prepare reports in a manner that is practical and meaningful for their specific services and business models including their global operations where those exists. In any case, Government should avoid striving to harmonise reporting across providers in an attempt to make data 'comparable' or, worse, to create any form of 'league table'.

It is also worth adding that the discussion surrounding the UK's Online Harms White Paper (including its draft transparency reporting template) is highly controversial and has not concluded.

With respect to the BOSE – and throughout the entire Discussion Paper – it appears that Government intends to quite substantially increase the powers of the eSafety Commissioner. Given the dominance of the online environment in the future, an enhanced role for the eSafety Commissioner may as such be appropriate. Such greater role, however, brings with it a need for robust checks and balances of these functions; something that is not articulated in the Discussion Paper. The Paper contemplates a power for the Commissioner to determine by legislative instrument the specific service providers to which the reporting obligations apply or to compel providers to supply reports on specific items.

Therefore, it will be important to clearly spell out the criteria that are to be satisfied and the oversight processes for such determinations. It may be helpful to analyse, and where feasible replicate, the powers and required processes of the Office of the Australian Information Commissioner (OAIC) and other Offices to develop a suitable model for the eSafety Commissioner.

It is not quite clear whether the BOSE would be presented to stakeholders for further consultation and, if so, at what stage. Will the BOSE be drafted in conjunction with the OSA and form part of a consultation on an exposure draft of the OSA, or will the development of the BOSE be a separate process that follows the OSA, noting that the OSA is only supposed to create the powers for the Minister to articulate a set of BOSE? Given the importance of the BOSE in the proposed new online safety framework, Industry would welcome clarification.

3.5. Broadening the cyberbullying scheme and establishment of a new cyber abuse scheme for adults

The Discussion Paper proposes to expand the existing cyberbullying scheme for children in two ways, i.e. through an increased scope of services (and providers), and a shortened response time from 48 to 24 hours. The Paper also considers additional tools for the Commissioner, such as a power to request or require that service providers enforce their terms and conditions in relation to a specific user.

The Paper also suggests the establishment of a new cyber abuse scheme for adults.

While the Discussion Paper does not expressly outline the scope of services to which the new adult cyber abuse scheme is proposed to apply, we assume that the range of services and providers covered is the same as for the expanded cyberbullying scheme.

Our comments below usually apply to both schemes, i.e. the cyberbullying scheme for children and the cyber abuse scheme for adults.

While the broadening of the scope of services may, prima facie, be appealing, it is not clear how the scheme would deal with some of the services now in scope in practice. For example, messaging services (e.g. WhatsApp, Signal, Telegram) are often end-to-end encrypted and do not offer an option for removal of individual parts of a conversation. Does this mean that user accounts would be required to be suspended, restricted or terminated when a complaint (that has been found valid) about cyberbullying of a child has been received?

In this context, it is also important to highlight that the consequences, i.e. the degree of harm that is likely to be incurred, are likely to be very different for content that is shared in a private messaging stream compared to the sharing of such content through public platforms accessible by a large number of individuals. In addition, private messaging services typically offer far greater controls and restrictions that enable the user to protect themselves from such harm.

Importantly, how would the eSafety Commissioner determine, in the context of a private communication between two individuals, whether a certain behaviour constitutes cyberbullying without extensive knowledge of the context and background of that communication? It would also be useful to consider to what extent harassment and menacing behaviours are already prohibited through existing statute.

Equally concerning is the broadening of scope to/inclusion of 'designated internet services' which basically includes any website.[4] Hosting services would also be included as indicated by the Discussion Paper.[5]

Many websites allow users to comment, post, chat or otherwise upload content. This includes product/service review websites, websites of clubs, schools, churches, social and charitable institutions etc. Sometimes, those content creation/upload functions require registration or the creation of a user account, other times these functions allow users to remain largely anonymous. Importantly, many of those websites are operated and maintained through very limited resources and/or volunteers. It appears unrealistic to expect the providers of such websites (or their hosts) to take-down content upon request within a 24 hour timeframe. We would think that many of these websites would struggle, even with far longer timeframes.

While it may be appealing to cast the web as widely as possible from a uniformity and enforcement perspective, we believe the proposed approach is simply not practical and a case for the inclusion of all kinds of services has not been made. The Discussion Paper fails to demonstrate the harm that emanates from such websites and that, where it is exists, it is proportionate to the proposed measures.

The expanded cyberbullying scheme and the adult cyber abuse scheme both appear to propose the inclusion of games, game streaming and game chat services into the scope of services that would be captured by those schemes. Many games indeed provide an internal chat function. However, without further evidence of the magnitude of the problem and, hence, any indication whether the proposed measures are proportionate, we are sceptical about the inclusion of those service. Similar problems as discussed above in the context of messaging services apply.

With respect to the contemplated powers for the Commissioner to request/require the enforcement of terms and conditions, it should be noted that providers themselves have very strong intention and incentive to enforce their terms. Where behaviour is evident on a digital platform that violates the platform's terms and conditions, this is often due to concerted efforts of users to evade those terms and limited abilities on the part of the provider to further pursue the enforcement. It is, therefore, not clear why the proposed powers are required: digital platforms would take action if they became aware of breaches of their terms and conditions. We are further concerned that exercising such a powers will result inconsistent application of a provider's terms and conditions, which are often global in nature.

Similarly, it is unclear, why a reduced timeframe for compliance with take-down requests is required. The Discussion Paper correctly notes that the existing regime is operating successfully and that requests to take down material have been met promptly (at times within 30 minutes) and with a 100 percent success rate. It appears unwarranted to shorten the timeframes and, at the same time, to expand the scope of services and providers captures under the scheme to include a wide variety of (often very small) services. Should Government proceed with reduction to a 24 hour time period, we believe there should be exceptions where an investigation requires more time to determine the nature and circumstances of the content or where consideration of an appeal from the party whose content is to be removed is required.

It is also worth asking if a civil penalties scheme for the adult cyber abuse scheme is warranted, given that the eSafety Commissioner has powers to issue civil penalties for image based abuse but, thus far, has not once used them, neither against services providers nor against individuals.

[4] *Enhancing Online Safety Act 2015*, Section 9A, Designated internet service: (1) For the purposes of this Act, *designated internet service* means: (a) a service that allows end-users to access material using an internet carriage service; or (b) a service that delivers material to persons having equipment appropriate for receiving that material, where the delivery of the service is by means of an internet carriage service;
[5] p.28, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019

It should also be noted that the statement that the German NetzDG requires removal of illegal content within 24 hours[6] is incorrect in its generality. The NetzDG only requires removal of 'manifestly unlawful' content within 24 hours. For content that is unlawful but not 'manifestly unlawful', providers have a seven-day deadline to remove or block access to the unlawful content. We also highlight that the NetzDG only applies to a much more limited set of services and providers, i.e. it applies to profit-making internet platforms that are intended to allow users to share content with other users or make it publicly available, but it exempts platforms offering their own editorial content. The exemption also extends to 'platforms intended for individual communication or the dissemination of specific content' (e.g. WhatsApp, Gmail). The law also exempts games, and providers who have fewer than two million registered users in Germany.

Importantly, the German NetzDG clearly confines the content that is subject to removal to illegal content, i.e. content that violates one (or more) of 21 statutes of the German Strafgesetzbuch (StGB) (Penal Code). While the requirement to assess content against these 21 statutes is not without its (serious) problems, it at least provides greater certainty and less room for arbitrary interpretation compared to a concept of 'harmful content' that is contemplated in other sections of the Discussion Paper or in the Charter. A comparison to this legislation without further details on its similarities and differences may, therefore, be misleading and ought to be avoided.

We also highlight that the NetzDG is highly controversial. The independent Research Services to the German Parliament (Wissenschaftliche Dienste des Deutschen Bundestages) have provided a report indicating concerns that the law may be unconstitutional and also violating European law.

3.6. Non-consensual sharing of intimate images (image-based abuse)

The Discussion Paper proposes a reduction for compliance with a removal notice of material subject to the image-based abuse regime from 48 hours to 24 hours.

As discussed in Section 3.5 above (*Broadening the cyberbullying scheme and establishment of a new cyber abuse scheme for adults*), we anticipate that a 24 hour timeframe would pose difficulties for many smaller operators of websites or content hosts. Consequently, we do not believe that the proposed reduction in timeframe is practical. A more graduated and nuanced approach to removal timeframes (and other aspects of the regime) might be needed to ensure that all providers can comply with a harmonised online content regime.

3.7. Addressing illegal and harmful content

We generally agree with the premise that Schedules 5 and 7 of the BSA and OCS contained in the associated industry codes are outdated and require urgent review.

It is key to highlight that Industry's ability to update codes to align with future technologies and changing community expectations relies on minimal prescriptiveness in the underlying legislation. The current codes that form part of the OCS are outdated and have not been revised due to the prescriptive nature of Schedules 5 and 7 of the BSA which prevented a meaningful overhaul of the codes. It is, therefore, of utmost importance to ensure that any revised underlying legislation, such as the proposed OSA, is sufficiently principles-based and flexible to allow future adjustments to the industry codes of a revised OSC.

Consequently, Industry welcomes the proposal made in the Discussion Paper to develop principles-based codes to better address content-related issues in an online world. We also agree with the understanding that such codes would need to be reflective of the nature and characteristics of the various online services providers and their services. Communications Alliance stands ready to work with Government, the Office of the eSafety

[6] p.28, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019

Commissioner, the ACMA and all relevant stakeholders to develop practical, fit-for-purpose codes.

With respect to the proposals directed at 'addressing harmful content', we feel that further detail and clarification are required in order to adequately assess the suggested path forward. In the following, we identify areas for clarification and further discussion.

It goes without saying that some material, such as child sexual abuse material, abhorrent violent material and content that promotes, incites or instructs in serious crime have no place on the internet and must be limited as far as possible. Indeed, our industry already cooperates closely and on a daily basis with the eSafety Commissioner and law enforcement agencies to combat such material online.

We have also engaged intensively with Government through the Taskforce processes in the aftermath of the Christchurch attack. ISPs have also voluntarily, and despite significant legal uncertainty, blocked access to websites that hosted the Christchurch attack material.

We are keen to explore with the relevant stakeholders how further progress to handle such online crises and serious crimes can be made.

Against this background, and also against the background of the Christchurch attacks in March 2019, it appears useful to provide the eSafety Commissioner with powers to assess such content instead of the content being classified by the National Classification Board. If the eSafety Commissioner is appropriately resourced, these powers should substantially improve the speed with which action can be taken and aid in reducing the spread of such material throughout the internet, thereby limiting the harm that emanates from it.

The additional powers to direct website blocking in certain circumstances will further assist with limiting harm in an online crisis. (Also refer to Section 3.9 of our submission, *Blocking measures for terrorist and extreme violent material online*.)

As such, we concur that some classes of content, such as child sexual abuse material, abhorrent violent material and content that promotes, incites or instructs in serious crime, ought to be considered 'seriously harmful content' and be dealt with it accordingly.

From a rule of law perspective, the boundaries of the definition of 'seriously harmful content' need to be carefully considered and ought to be necessary and proportionate to legitimate aims. The Discussion Paper seems to suggest that there might be (substantially?) more content that could be deemed seriously harmful and that the three stated categories may only form examples of such content. In other words, it will be key to understand which other provisions, if any, Government is envisaging to include into the definition of seriously harmful content.

Depending on the breadth of content that would be deemed seriously harmful, the additional powers to assess such content and to issue take-down notices accordingly, risk moving jurisdictional and enforcement powers from the Courts to the eSafety Commissioner – something that is, independent of the person and office itself, not desirable in a democratic society.

Similarly, the Discussion Paper proposes a power for the Minister to make a legislative instrument to capture additional types of content that may, but are not necessarily, criminalised under the Commonwealth Criminal Code (Criminal Code). This creates additional risks that the treatment of specific online content could be subject to political discretion. It would be more appropriate to limit content classified as seriously harmful to a small number of statutes (as per Criminal Code) and to amend the Criminal Code if it was felt that certain behaviour/material was sufficiently harmful, and indeed criminal, that it would warrant coming under this definition. It could be argued that any material that would be subject to measures that restrict freedom of speech, would be of such serious nature that its treatment requires consideration in the Criminal Code anyway.

We note that our assessment of the definition and treatment of seriously harmful content is based on the understanding that any such content would be subject to a notice-and-take-

down approach, rather than a requirement to detect (or be aware of) and remove such content proactively. (Figure 6 of the Discussion Paper[7] seems to suggest that the former is the case.) While large platforms and search engines go to great lengths to proactively detect and remove such material, this approach (a monitoring and detection requirement) would not be appropriate or practical for all industry participants and would require further discussion.

Unfortunately, it is not clear whether a notice-and-take-down approach would also be envisaged for prohibited content. The Discussion Paper states "In contrast, designated internet services, *such as websites, have a greater degree of control over the content made available through their services*. To this end, it is expected that the codes applicable to these types of services would require that content that would otherwise be classified RC or X18+ must not be hosted in Australia. Breaching a code provision relating to the *hosting* of X18+ or RC would be treated very seriously and could trigger an investigation by the eSafety Commissioner into the *content host*."[8] (Emphasis added.)

From the above it is not clear if the expectation is that the codes would link to a notice-and-take-down regime, with notices being issues by the eSafety Commissioner, or whether the hosting in itself would already constitute a code breach and subsequent investigation and enforcement.

Against this background it is also not clear whether the Paper has sought to make a distinction between website operators, who indeed have a degree of knowledge of and control over the content that is on their websites, and hosting providers who merely provide the infrastructure for content and have no visibility of the content that they host and, consequently, no ability to remove it unless expressly notified. It should also be understood that the ability of a hosting provider to take down content on a granular level can be very limited and content removal (where the supplier of the content does not cooperate with the hosting provider) may imply the removal of entire accounts and/or services. (The current industry code reflect this by only requiring content hosts to educate content providers about their legal obligations (and such education can be contained in contractual documents), to put in place appropriate access controls for restricted content, and to comply with take-down notices.)

The same confusion and lack of differentiation between different types of services arises with respect to seriously harmful content. The Discussion Paper indicates that the take-down requirements for such content would apply to social media services, designated internet services and relevant electronic services[9] (it does not list hosting services) but mentions a harmonisation with other regimes which do apply to hosting services.

Any expectation on service providers that explicitly or implicitly requires them to monitor content (and potentially to remove content) has to be treated with caution as not all providers have the ability to do so and/or an expectation to monitor content ought not be placed on them. To the extent that some form of content moderation can be expected of some types of providers, it is key to strike an appropriate balance between the desire to limit the occurrence of the worst types of illegal content on the internet and freedom of speech that all individuals ought to be afforded in democracies. This balance will be significantly influenced by a variety of factors, including:

- the extent to which those providers are required to apply their own judgement to online content – this may be influenced by the question whether content is clearly defined as illegal or whether content is deemed harmful with less clear boundaries as to the definition of harm;
- technical capabilities – most likely in the form of algorithms for content recognition and categorisation;
- the timeframes for removal of content; and

[7] p.40, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019
[8] p.37, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019
[9] p.39, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019

- penalties for non-compliance with any removal obligations.

All of these factors will play a crucial role in determining the percentage of successfully removed illegal content and the degree of over-blocking that may result from content moderation/removal requirements that are poorly targeted and/or difficult to comply with.

Consequently, given the importance of the proposals in a new OSA, we urge the Department to clearly articulate what is being envisaged with respect to the definition, detection and treatment of prohibited, seriously harmful and harmful content for the different types of service providers.

It would also be helpful to get an understanding if 'inappropriate content' (also used in this context) is being used as a synonym for harmful content or constitutes a different concept which would require additional explanation and very careful consideration.

The Discussion Paper also contemplates a proposal to require all sectors of industry "to provide their users with access to the best available technology solutions to help Australian families to limit access to *prohibited* content […]. The kinds of tools that could be used under the codes are explored further in the section on opt-in tools and services to restrict access to *inappropriate* content. The concept of *harmful* content under the codes would be informed by the National Classification Code, and the technology solutions deployed would be proportionate to the potential harm posed by the material."[10] (Emphasis added.)

As indicated above, the use of different and unclear terms makes it hard to understand (and comment on) what is expected of Industry: is the use of best available technology directed at prohibited content, i.e. RC and X18+, or at harmful content, which we take to mean MA15+ and R18+ (Figure 6), or at inappropriate content, which has not been defined in the Paper?

Furthermore, unfortunately neither the section *Addressing illegal and harmful content online* nor the section *Opt-in tools and services to restrict access to inappropriate content* (note again the use of the term 'inappropriate') of the Discussion Paper include detail on the defining factors of 'best available technology'.

The use of this term and the associated concept are impractical for a number of reasons, including:

- Who would be the judge of what technology is best? What would qualify such a person/authority to make this judgement? (E.g. the photo recognition tool of company A may be picking up a larger number of illegal photos subject to its algorithm but is not robust for edited pictures. The tool used by company B detects edited photos but has a narrower range – which one is better or even best?)
- 'Best' for whom? The provider or the end-user?
- 'Best' for which type of service or service provider?
- What constitutes availability? Availability at any price/time? How would this idea sit with proprietary solutions?
- At what point in time are 'best' and 'available' being assessed? Would providers need to constantly update their technology and even completely change technologies to reflect what is deemed best?
- Is such technology to be provider-based or could it be end-user implemented?

Consequently, we suggest amending the concept to something along the lines of a requirement to make available 'appropriately effective technology', and to strengthen such a concept with robust guidance to assist providers understanding the expectations that are being placed on them.

With respect to the suggested 24 hour take-down timeframe for seriously harmful content, we reiterate our concerns that this timeframe is likely to be impractical for many providers, no matter how desirable it may be to remove such content with utmost urgency.

[10] p.37, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019

The Discussion Paper also proposes to extend the take-down regime to content hosted overseas. While it is certainly desirable that content be treated equally independent of its location, it is likely that enforcement of take-down requests in overseas locations, if met with non-compliance, would be difficult.

3.8. Opt-in tools and services to restrict access to inappropriate content

Our industry supports individuals who seek to limit content that they or their families can access on the internet. As indicated in the Discussion Paper, a variety of approaches, tools and services exist to suit different types of services and needs. The accreditation of Family Friendly Filters (FFFs) is one such approach.

Unfortunately, it is also the case that many people (the majority?) do not or only sparingly make use of the tools already available to them despite surveys that have identified online safety as a major concern.

Consequently, we are very keen to engage with all stakeholders to examine ways in which users can be better motivated and enabled to access existing and new tools that meet their needs in a timely manner. This could include default safety settings at a device level as well as easy ways to implement filters at a user level and other measures that are offered, where appropriate and applicable, at key touch-points during the sales and activation process across the supply chain for online services, and during the service provision life-cycle.

We are unclear if the Discussion Paper or Government policy contemplate network-based filtering. If so, we would object to any requirements that would amount to filtering at a network level. Such filtering would be fraught with difficulties with regards to ensuring that different user groups have the appropriate level of access to content. This technology would also be a significant regulatory burden and could place insurmountable technical/financial difficulties for smaller ISPs. A suite of alternative arrangements and powers, as proposed in the Discussion Paper, constitutes a more proportionate, effective and practical approach to ensuring that access to specific online content is limited or removed to the extent possible.

We welcome the suggestion to create industry codes to require service providers to better enable and protect users from certain types of content, and we are willing to engage with all stakeholders on this issue.

As in the previous section, we raise concerns around the use of the term 'best available technology' and would welcome further discussion around a more practical concept.

We are also happy to explore whether already existing information provision requirements need updating or supplementing, noting that users often already exhibit 'information fatigue' and any information provision needs to be sensitive to the context within which it is provided. Service providers also require sufficient flexibility to provide information in the most effective manner, given the existing information provision requirements, varying service delivery models, user base and services, i.e. any prescriptive approach to information provision ought to be avoided.

However, it will be critically important that any such efforts by industry to share online safety information are bolstered with a substantive Government-led and funded awareness campaign similar to the 'Slip, slop, slap' or the digital TV switch-over campaigns. Given Government and Industry efforts to also improve cyber security awareness among Australians and small businesses, such a campaign, if well executed, might also provide an opportunity to include a simple key message around the importance of updating default and creating strong passwords for user accounts etc.

The Discussion Paper proposes an accreditation scheme, either run by the eSafety Commissioner or by an industry body, such as Communications Alliance, to promote a baseline of accessible and affordable opt-in tools and services. The Paper contemplates that "[t]his could be achieved through an accreditation program to evaluate the mainstream tools and services available in the market." This suggests that, beyond an accreditation of

filtering tools (including FFFs), it might be envisaged to accredit a variety of tools from a range of types of services providers.

We would welcome further discussion as to how such an accreditation program could be designed so that it overcomes the issues identified in the Discussion Paper (knowledge gaps, sustainability, lack of information about good practices) without becoming overly complex and, thereby, limiting its effectiveness. For example, would such a program simply distinguish between accredited and non-accredited tools, which may provide easy to digest information but may give limited guidance as to whether a respective tool suits a specific user group or need, or should this program consider some form of star rating, or would different categories of accreditations be most valuable for users?

The resources required for the development and ongoing delivery of such a program are likely to be substantial and, as we understand it, neither the eSafety Commissioner, and certainly not Communications Alliance are currently equipped or funded to deliver an accreditation program.

The accreditation program currently run for the FFFs is provided through a commercial testing lab which receives fees for its services from the filter providers that wish to have their filters (re-)certified.

3.9. Blocking measures for terrorist and extreme violent material online

The Discussion Paper proposes a new, dedicated power for the eSafety Commissioner to direct ISPs to block domains containing terrorist and/or extreme violent material in an online crisis event.

It is important to understand that ISPs did not receive such a direction (despite desperate attempts by ISPs to be issued with a direction), from the eSafety Commissioner or from any other authority, during and in the aftermath of the Christchurch attacks. In the absence of a direction to block the websites that hosted the footage of the shootings and the manifesto, all major Australian ISPs took the decision to block the identified websites 'off their own bat'. This left them exposed to legal liability – a situation which lasted almost 6 months!

We have since worked with the eSafety Commissioner to put in place a protocol that governs the processes during online crisis events. The proposed direction power would complement this protocol, and we welcome the proposal in principle.

However, it is not quite clear from the Discussion Paper what criteria, if any, the eSafety Commissioner would apply to determine whether such an event is evolving and what would constitute 'terrorist and extreme violent material'? From a rule of law perspective, this uncertainty is very concerning. Is it envisaged to link this assessment to statutes within the Criminal Code or would the assessment be open to the Commissioner's discretion? It seems that such material would constitute a subset of what the Paper describes as 'seriously harmful content' which in turn is envisaged to be defined by reference to the Criminal Code. We note that the recommendations developed through the Taskforce to Combat Terrorist and Extreme Violent Material Online already contained a definition of such material, and we suggest using this definition in the OSA.

We also note that not only the provisions around civil immunity of the *Guidelines for the use of section 313(3) of the Telecommunications Act 1997 by government agencies for the lawful disruption of access to online services* ought to be mirrored but <u>all</u> aspects of these Guidelines, including the requirement to have access to appropriate technical expertise (this is not casting judgement on the Office's current technical expertise!).

We also welcome a move towards automation of website blocking requests, not only for online crisis events but more broadly for blocking requests made under other legislation, e.g. the *Copyright Act 1968* and Section 313(3) of the *Telecommunications Act 1997*. However, it may be premature to mandate automation. This would require further analysis of the

technical capabilities of ISPs and the timeframes involved to potentially develop solutions that would be able to receive and implement requests in an automated manner.

3.10. Ancillary service provider notice scheme

We note that, at this stage, Communications Alliance does not represent the online gaming industry.

In principle, we do not oppose an ancillary service provider notice scheme which would create an avenue for the eSafety Commissioner to request (not require) the delisting/de-ranking and the cessation of distribution of certain content.

It is also welcome that the proposal intends to use these powers as reserve powers which would only come into play if more direct approaches of the primary providers of the content have not been successful.

However, we are concerned with the practical application of the powers and the difficulty of striking a balance between limiting harm and providing a service that may be of great value to a large number of users.

Under the proposal, it is envisaged that the service must be "systemically and repeatedly facilitating the posting"[11] of certain categories of material. How would this threshold be applied to, for example, a very popular online game (e.g. Fortnite) whose in-game chat function might be used to 'repeatedly and systemically' bully one or a number of players. How would this abuse and the powers to request a delisting or to cease distribution be balanced against the enjoyment that the game brings to millions of other gamers who do not engage in any objectionable behaviour?

Against this background, we recommend limiting the new powers to services whose primary or predominant purpose or actual use is to facilitate cyber bullying, cyber abuse, image-based abuse or the dissemination of seriously harmful content. This would help with the stipulated aim to ensure that the scheme is "proportionate and appropriate to the specific roles that these services play in the online ecosystem."[12]

3.11. Role of the eSafety Commissioner

We note the discussion on the future governance arrangements of the eSafety Commissioner. While we have not formed a final opinion on this issue and noting that our immediate focus rests on assisting with the development of a fit-for-purpose new OSA, we tentatively submit that the Office of the eSafety Commissioner, being an independent statutory Office, appears to be well placed within the broader communications and media remit of the Australian Communications and Media Authority (ACMA).

At this stage, we do not see merit in the creation of a separate entity. It is also difficult to envisage the Office functioning efficiently within the already large Department of Infrastructure, Transport, Regional Development and Communications, or another department.

[11] p.47, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019
[12] p.47, *Online Safety Legislation Reform – Discussion Paper*, Department of Communications and the Arts, Dec. 2019

4. CONCLUSION

We look forward to further engaging with Government, the eSafety Commissioner and other stakeholders on the mutual desire to ensure that the Australian community is well-equipped to safely enjoy online environments.

We welcome the proposal for Industry to create principles-based codes to deal with key aspects of the new online safety regime, and we stand ready to engage with all stakeholders to facilitate the required code development processes.

We are also keen to closely cooperate with Government to develop an education and awareness campaign for the Australian public to ensure that end-users are empowered and motivated to protect themselves, as far as possible, from online harms and practice responsible online behaviours.

Noting a number of issues that require further clarification and discussion, we welcome an ongoing dialogue with key stakeholders.

For any questions relating to this submission please contact Christiane Gillespie-Jones on 02 9959 9118 or at c.gillespiejones@commsalliance.com.au.

Published by:
COMMUNICATIONS
ALLIANCE LTD

Level 12
75 Miller Street
North Sydney
NSW 2060 Australia

Correspondence
PO Box 444
Milsons Point
NSW 1565

T 61 2 9959 9111
F 61 2 9954 6136
E info@commsalliance.com.au
www.commsalliance.com.au
ABN 56 078 026 507