Consultation on a Bill for new Online Safety Act

GLOBAL PARTNERS DIGITAL

Global Partners Digital submission

February 2021

About Global Partners Digital

Global Partners Digital is a social purpose company dedicated to fostering a digital environment underpinned by human rights.

Introduction

We welcome the opportunity to provide feedback on the exposure draft Online Safety Bill to improve Australia's online safety legislation. GPD recognises the legitimate desire of the Australian government to tackle unlawful and harmful content online, and the majority of the proposals put forward in the draft Online Safety Bill are reasonable and sensible. We are also pleased that certain elements of this Bill appear to reflect a number of our recommendations made in our previous consultation response in 2020. Based on our analysis, however, we believe that particular aspects of the Bill, if taken forward in their current form, may still pose risks to individuals' right to freedom of expression and privacy online and could be inconsistent with Australia's international human rights obligations.

In this response, we relay our concerns and make a series of recommendations on how the proposals could be revised to mitigate these risks. We believe these considerations and recommendations, if incorporated into the final legislation, will help safeguard freedom of expression and privacy online.

Framework for analysis of the draft Online Safety Bill

Our analysis of the draft Online Safety Bill is based on international human rights law, specifically the International Covenant on Civil and Political Rights (ICCPR). Article 19 of the ICCPR guarantees the right to freedom of expression, including the right to receive and impart information and ideas of all kinds regardless of frontiers. Article 17 of the ICCPR guarantees the right to privacy and provides that "no one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence". Restrictions on the right to freedom of expression or privacy guaranteed under international human rights law are only permissible when they can be justified. In order to be justified, restrictions must meet a three-part test, namely that: (1) restrictions are provided by law; (2) restrictions pursue one of the purposes set out in Article 19(3) of the ICCPR - to protect the rights or reputations of others, to protect national security or public order, or public health or morals; and (3) restrictions must be necessary and proportionate, which requires that the restriction be the least restrictive means required to achieve the purported aim.

It is important to remember that Australia's obligation to ensure that these rights are not unjustifiably restricted exists both in relation to restrictions which stem from the actions of the state itself as well as those caused by third parties, such as private companies. As such, it makes no difference from the perspective of the individual affected whether any restrictions are imposed and enforced directly by the state (e.g. through creating criminal offences which are

enforced by the police and the courts) or through third parties, particularly when the third party is acting in order to comply with legal obligations.

Please see our original 2020 submission to the consultation on online safety reforms for a more detailed examination of this methodology.

Human rights analysis of the draft Online Safety Bill

Overarching Elements

We welcome the inclusion of Section 233(1) of the draft Bill, which states "This Act does not apply to the extent (if any) that it would infringe any constitutional doctrine of implied freedom of political communication". This reference is critical considering the potential negative impacts on freedom of expression posed by the proposed legislation, but it fails to ensure protection for the full right to freedom of expression under international law. Australia's constitutional protections, including the implied freedom of political communication, fall far short of the right to freedom of expression under international human rights law. In our original 2020 submission, we noted that the principle of "balancing the competing objective of user safety and freedom of expression" be modified to "uphold the right to freedom of expression". While Section 233(1) appears to partly reflect this recommendation, the text could be improved with additional reference to Australia's international human rights obligations. This would be particularly useful considering that Australia has limited constitutional protections for freedom of expression.

Recommendation 1: We recommend that Section 233(1) be amended to include explicit reference to Australia's obligations under the International Covenant on Civil and Political Rights. For example, "This Act does not apply to the extent (if any) that it would infringe any constitutional doctrine of implied freedom of political communication or the right to freedom of expression under Article 19 of the International Covenant on Civil and Political Rights".

We are pleased that Section 24(1) of the draft Bill contains a stand-alone reference to the Convention on the Rights of the Child and provides that the Commissioner must, as appropriate, have regard to the Convention in the performance of functions under the Act and in relation to Australian children. The Convention on the Rights of the Child sets out a broad range of rights that all children are entitled to. Section 24 would therefore require the Commissioner to have regard to the full array of rights provided for in the Convention, including a child's right to freedom of expression (Article 13) and the right to privacy (Article 16). Still, it would be beneficial for this section to include a more clear-cut recognition of the full range of rights encompassed under the Convention.

Recommendation 2: We recommend that Section 24(1) be modified to include explicit reference to the full range of rights provided for under the Convention on the Rights of the Child. For example, "The Commissioner must have regard to all civil, political, economic, social and cultural rights enumerated in the Convention on the Rights of the Child in the performance of functions: (a) conferred by or under this Act; and (b) in relation to Australian children".

There should be a similar duty on the Commissioner to consider the International Covenant on Civil and Political Rights. This is because there are a number of functions and duties of the Commissioner that will engage the rights protected by the ICCPR. We suggest that the draft Bill contain a stand-alone reference to the ICCPR, which requires the Commissioner to have regard to the Covenant in the performance of functions under the Act.

Recommendation 3: We recommend that the draft Bill contain a stand-alone reference to the International Covenant on Civil and Political Rights, which requires the Commissioner to have regard to the Covenant, including the rights to freedom of expression and privacy, in the performance of functions under the Act. For example, amending Section 24 to include a new subsection: "(2) The Commissioner must, as appropriate, have regard to the International Covenant on Civil and Political Rights in the performance of functions: (a) conferred by or under this Act; and (b) in relation to Australians."

Basic Online Safety Expectations (BOSE)

We welcome that Section 47 of the draft Bill would require the Minister to consult with the public and invite interested parties to submit written comments when making a BOSE determination for a particular service or class of services. We are also pleased that the Minister must have due regard to any comments submitted when making a determination. But it would be preferable if the Minister was required to consider specific factors when making a Section 45 determination, including that they make determinations as narrowly as possible. This will help ensure that companies are only required to satisfy the Basic Online Safety Expectations where there is clear evidence or risk of harm. The scope of companies required to adhere to these expectations should be proportionate to the type of service and the amount of harm that takes place on a respective service. Blanket decisions on whole classes of services, as described in Section 45, are unlikely to constitute a narrowly tailored and proportionate response.

Recommendation 4: We recommend that the draft Bill require the Minister to make BOSE determinations on which companies are within scope as narrowly as possible. The scope of companies required to adhere to these expectations should be proportionate to the type of service and the amount of harm that takes place on a respective service.

Furthermore, we believe there is insufficient detail on particular elements of the core expectations outlined in Section 46 that may pose potential risks to individuals' freedom of expression and privacy without additional clarification. Section 46 indicates that service providers must take 'reasonable steps' to ensure that end-users are able to use a service in a safe manner. There is an expectation that, in determining what are such 'reasonable steps', that the provider will consult with the Commissioner. However, there is no indication that the Commissioner is required to provide detailed guidance on what actions or processes satisfy core expectations and recommend this be added within Section 46.

We are pleased that one of the core expectations set out in Section 46 will require services to provide clear and readily identifiable mechanisms that enable end-users to report and make complaints about various forms of illegal and harmful material. But given the wide range of content prohibited in the proposed framework, it would be beneficial for services to be able to consult with the Commissioner to ensure that their decisions do not impermissibly restrict users' right to freedom of expression.

Recommendation 5: Companies that are required to adhere to the BOSE and take 'reasonable steps' set out in the core expectations should be provided an opportunity to request further guidance from the Commissioner where they reasonably believe that upholding the core expectations might undermine their ability to safeguard freedom of expression or privacy.

Services must also take 'reasonable steps' to minimise the extent to which cyber-bullying material, cyber-abuse material, non-consensual intimate images, class 1 material, and abhorrent violent material are available on their services. These core expectations could encourage

proactive monitoring of content and the unintentional removal of permissible content. Given the scale of content which is generated and shared online, companies will increasingly turn to automated processes, including AI, to meet their obligations. The risk here is that automated processes will detect and remove content that is not actually unlawful or harmful in a particular context. Automated processes have had some success in relation to content moderation with types of images, including the ability to identify copies of images that have already identified by humans as constituting child sexual abuse and exploitation. However, automated processing has been less effective when identifying speech or less specific forms of unlawful or harmful content, such as cyber-bullying material or cyber-abuse material.

Recommendation 6: Section 46 should clearly indicate that taking 'reasonable steps' to minimise illegal and harmful forms of content does not require a service to use automated processes to proactively monitor and remove content. If automated decision-making is undertaken to meet core expectations, this should be accompanied by requirements to ensure the use of open source tools, transparency around standards, and appropriate appeals mechanisms.

We are particularly concerned that the core expectations may pose risks to encryption and individuals' right to privacy. While the draft Bill does not reference encrypted services, any requirement in the BOSE to filter or monitor material which applied to encrypted and other private channels would almost certainly amount to an unjustifiable restriction on individuals' right to communicate privately. This is because such services would need to remove or weaken privacy-enhancing technologies, such as encryption, in order to be able to filter or monitor material content which is generated or shared using them. We therefore suggest that the Bill explicitly note that companies are not required to cease, restrict or in any way weaken their use of encryption or other privacy-enhancing technologies to satisfy core expectations.

Recommendation 7: Section 46 should clearly indicate that companies 'reasonable steps' to satisfy core expectations do not include the filtering or monitoring of material, if they would require a service to restrict or in any way weaken their use of encryption or other privacy-enhancing technologies.

Section 46 indicates that designated services will have to take 'reasonable steps' to ensure that technological or other measures are in effect to prevent access by children to class 2 material. The lack of clarity around 'reasonable steps' poses a potential risk to individuals right to privacy. As noted in our 2020 submission, there are particular technologies that could be used to satisfy this expectation, such as facial recognition technology, but these types of technological solutions involve the processing of large amounts of data, often personal data, when employed for identification, profiling or age verification. We recommend that service providers are not required to employ any form of technology that may pose risks to individuals' right to privacy in order to satisfy the 'reasonable steps' element of this core expectation.

Recommendation 8: Section 46 should clearly indicate that taking 'reasonable steps' to ensure that technological or other measures are in effect to prevent access by children to class 2 material does not require service providers to use technologies that pose risks to freedom of expression or privacy, such as facial recognition technologies. If these technological measures are to be pursued they should be accompanied by sufficient safeguards, including comprehensive data protection measures being taken by those who collect or process any personal data, and oversight by a competent authority or regulatory body.

The draft Bill would empower the Commissioner to establish reporting requirements under the BOSE, which would mandate services report on their compliance with one or more of the specified expectations. It would, however, be beneficial if these reports also contained relevant information on how freedom of expression and privacy were protected by particular services. This type of transparency requirement has been proposed in the UK Online Harms White Paper, and the recently released full government response indicates that "Certain companies will also need to produce transparency reports, which are likely to include information about their measures to uphold freedom of expression and privacy". These reports will include "information about the measures and safeguards in place to uphold and protect fundamental rights, ensuring decisions to remove content, blocking and/or delete accounts are well founded". This approach would be considered as best practice and should be emulated in the draft Bill.

Recommendation 9: We suggest that, as proposed in the UK Online Harms White Paper, that reporting requirements include information on how companies are respecting freedom of expression and privacy on their services.

Takedown Schemes

We continue to be concerned over the lack of adequate appeals mechanisms for removal and blocking notices made by the eSafety Commissioner for private companies and end-users. International human rights law requires that any person whose rights or freedoms are violated shall have an effective remedy, which is guaranteed under Article 2(3) of the ICCPR. We appreciate that the draft Bill does not propose making private entities decide whether a particular piece of content is lawful or unlawful, and welcome that Section 220 would provide private companies and end-users the ability to challenge decisions in the Administrative Appeals Tribunal. Still, while the decision making of a public body can provide a far greater level of transparency and accountability, additional opportunities to challenge take-down notices or other types of decisions should be provided for within the proposed framework. These additional appeals mechanisms, specifically those between the Commissioner and particular companies or end-users, would be beneficial because civil proceedings and other forms of redress are often cumbersome, time-intensive and expensive. Meaningful opportunities to challenge decisions should be readily available and accessible to the public before resorting to the courts.

Recommendation 10: The proposed takedown and blocking schemes should enable all end-users and private companies the opportunity to challenge decisions made by the Commissioner before resorting to the court system. The Commissioner should have the resources available to provide an effective remedy, which should include the ability for content to be reinstated.

We continue to be concerned that private companies will be required to take action in a reduced time frame upon receiving a removal notice from the eSafety Commissioner. We understand the need to quickly respond to take-down notices for image-based abuse, cyber abuse, cyber-bullying and seriously harmful content. But we believe that the shorter 24 hour period may not be practical for certain companies, particularly smaller companies or newer types of relevant electronic services, such as online gaming services, which are not included in the scope of the existing framework. Many of these smaller companies and newer services will not have the capacity or dedicated structures to respond in such a short time-frame.

[1] Online Harms White Paper: Full government response to the consultation (Dec 2020), available at: https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response

Recommendation 11: Smaller companies and newer services should be provided with a more flexible time frame when they are unable to comply with the 24 hour take-down requirement. They should also be able to seek assistance from the Commissioner if they are unable to develop the necessary internal structures to be able to respond to notices.

- *Cyber-bullying Scheme*

We are pleased that end-users will not be faced with civil fines for lack of compliance with removal notices under this scheme and that only private companies would be subject to civil penalties. This approach recognises that the recipients of such end-user notices are likely to be children themselves, and that this scheme pertains to harmful, but not necessarily 'illegal' forms of content. We are, however, concerned that the eSafety Commissioner would seemingly have the ability to pursue an injunction against a child for failing to comply with an end-user notice without first submitting a formal warning to the end-user or before sending a removal notice to the relevant service. These changes would reflect a more proportionate response to enforcement.

Recommendation 12: We recommend that the Commissioner be required to pursue alternative courses of action to enforce removal notices sent to end-users under the cyber-bullying scheme before resorting to injunctions. The Commissioner should be required to either issue a warning notice to a particular end-user that fails to comply with a removal notice, or they should be required to send a removal notice to the relevant service.

- *Adult Cyber-abuse Scheme*

We are pleased that the Adult cyber-abuse Scheme would align with the threshold established in the Criminal Code Act 1995, which we see as an appropriate means of tackling the issue. This type of approach is beneficial as it avoids the creation of two separate legal regimes for different domains - one for the online environment and another for the offline environment.

- *Image-based Abuse Scheme*

We welcome that Section 86 of the draft Bill provides exemptions for end-users of social media services, relevant electronic services or designated internet services who post an intimate image of a person. According to this section, a post may be exempt if it, amongst other reasons, is for genuine medical or scientific purpose, or if an ordinary reasonable person would consider the post as acceptable, having regard to the nature, content and circumstances of the post and the age, intellectual capacity, vulnerability of the depict person, and the degree to which the post of the intimate image affects the privacy of the depicted person. This nuanced exemption recognises the need for a contextual examination of content, which ultimately will reduce potential risks to freedom of expression. We are especially pleased that the exemption makes specific reference to the privacy of the depicted person and suggest that this language be retained in the final version of the Bill.

Recommendation 13: We recommend that the exemptions provided for in Section 86 be included within the final Bill as they help mitigate risks to freedom of expression and privacy.

- *Online Content Scheme*

We are particularly concerned about the reduced 24 hour time-frame under this scheme as the Commissioner would be able to issue removal notices for class 1 material to service providers based in Australia or abroad. The global reach of this power could pose risks to individuals' right to freedom of expression as it assumes that all services will be able to geo-block particular

material in Australia, as opposed to simply resorting to the global removal of content. We suggest that additional safeguards be built into the Bill, as seen in Recommendation 3, to require the Commissioner to consider the impact that a particular removal notice might have on freedom of expression when sent to a service based outside the country.

We are concerned that the threshold for issuing link deletion notices and app removal notices is too low and does not account for the varying sizes of services within scope. Section 124 provides that the Commissioner may only give a link deletion notice when they are satisfied that there were 2 or more times during the previous 12 months when end-users in Australia could access class 1 material using a link provided by the service, and the Commissioner gave one or more removal notices in relation to class 1 material that were not complied with. The same threshold is provided with regard to app removal notices in Section 128. But there is a high likelihood that larger services may inadvertently satisfy these conditions given their broad usage with millions of end-users and the scale of removal notices received. Failure to comply with a singular removal notice relating to class 1 material could still trigger a link deletion notice or app removal notice even if the vast majority of removal notices were complied with. We disagree that failure to comply with one or more removal notices relating to class 1 material during a 12 month period would constitute "systematically" ignoring take down notices as is noted within the 'Online Safety Bill - Reading Guide'. We suggest this criterion be reevaluated and that a more flexible threshold be established which accounts for all types of services within scope.

Recommendation 14: We recommend that the threshold for issuing link deletion notices and app removal notices under the Online Content Scheme be reevaluated to account for the varying types of services within scope. We suggest that the text of Sections 124(4) and 128(4) establish a more flexible threshold that would allow for notices to only be issued when a website or app was determined to truly 'systematically ignore' take down notices for class 1 material.

Abhorrent Violent Material Blocking Scheme

We welcome that Section 104 of the draft Bill provides exemptions for particular types of content that would otherwise be subject to the Abhorrent Violent Material Blocking Scheme. We are particularly pleased that the exemption applies to material when needed for scientific, medical, academic, public interest purposes, amongst others. In our 2020 submission, we suggested that exceptions be made for specific types of material as seen in Section 474.37 of the Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019. The exemption included in the draft Bill is closely aligned with the one provided for in this 2019 law and we recommend that this be retained in the final framework.

Recommendation 15: We recommend that the exemptions provided for in Section 104 are retained in the final version of the Bill.

We also welcome that there is a clear requirement in Sections 95 and 99 for the Commissioner to be satisfied that the availability of AVM material online is likely to cause significant harm before issuing a blocking request or notice. The Commissioner must have regard to the nature of the material and number of end-users who are likely to access the material. The Commissioner must also consider whether other powers exist to accomplish the same objective and any other matters as the Commissioner considers relevant. These restrictions which limit the Commissioner's power to issue blocking requests or notices reflect a proportionate approach.

Recommendation 16: We are pleased with the limitations and exemptions outlined under this scheme, specifically in Sections 95 and 99, and recommend they are included within the final

version of the Bill. We believe it is especially important that the Commissioner have regard to these limitations, and consider the impact on freedom of expression across all functions under the Act, as noted in previous recommendations.

As noted above, we strongly advise that adequate appeals mechanisms be provided for end-users and private companies to contest removal and blocking decisions made by the Commissioner across all schemes. But it is especially important that appeals mechanisms are available to ISPs under this scheme as a blocking requirement may have widespread and disproportionate impact on freedom of expression. Section 100 specifies that the duration of a blocking notice may last no longer than 3 months, but it then indicates that the Commissioner would be able to issue a fresh blocking notice that comes into force immediately after the expiry of the original blocking notice. We are concerned that further safeguards or limitations are not built into this Section as they relate to the renewal of blocking notices.

Recommendation 17: We strongly recommend that adequate appeals mechanisms be provided for ISPs to challenge blocking notices issued by the Commissioner under the Abhorrent Violent Material Blocking Scheme, and request that ISPs be given the opportunity to challenge any extension or issuing of a fresh blocking notice.

Information Gathering Powers

We have some concern that the information gathering powers provided for in Part 13 of the draft Bill may pose a potential risk to individuals' right to privacy without additional clarification on the scope of these powers. Section 194 empowers the Commissioner to obtain the identity or contact details of an end-user from a person when they are the provider of a social media service, relevant electronic service, or designated internet service. We understand that the Commissioner must have sufficient information gathering powers to effectively carry out its functions, but are nonetheless concerned that the threshold for issuing a written notice to a particular provider is relatively low. The Commissioner only needs to "believe on reasonable grounds" that the information is relevant to the operation of the Act. This threshold is inconsistent with the language used on page 12 of the 'Online Safety Bill - Reading Guide', which instead notes that this power will only be used to obtain contact details or the identity of an end-user "if necessary". We suggest that this higher threshold be incorporated into the Bill.

Recommendation 18: Section 194 of the draft Bill should be modified and replaced with the language provided for in the Reading Guide. Specifically, the threshold required for issuing a written notice should be whether the Commissioner determines such information to be "necessary to the operation of this Act", as opposed to when the Commissioner simply believes on reasonable grounds that the information is relevant.

Moreover, Section 195 requires that the provider of a social media service, relevant electronic service, or designated internet service comply with written notices "to the extent that the person is capable of doing so" or they could face a substantial fine of 100 penalty units. It is unclear what is meant by "to the extent that the person is capable of doing so" and whether services which use end-to-end encryption would fall within the scope of this exception.

Recommendation 19: We recommend that Section 195 clearly indicate that a person does not need to comply with a written notice under section 194 to the extent that it would require the provider to decrypt encrypted communications, or to cease, restrict or in any way weaken their use of encryption.