Draft Online Safety (Basic Online Safety Expectations) Determination – Submission to the Department of Infrastructure, Transport, Regional Development and Communications

RE: Consultation on the draft Online Safety (Basic Online Safety Expectations) Determination

To Whom It May Concern: Thank you for the opportunity to provide this submission to the Department of Infrastructure, Transport, Regional Development and Communications consultation into the draft Online Safety (Basic Online Safety Expectations) Determination ("Determination").

The Online Safety Act 2021 ("Act") aims to improve and promote online safety of Australian's by placing greater responsibility on service providers to ensure their services provide a safe environment for Australian users. I, like many others recognize this is an important part of building a digitally inclusive landscape, however, many aspects of the draft Expectations, if taken forward pose substantial risk to individuals security and privacy online.

Defining Online Safety

During the consultation period of the Act in February 2021, I provided a submission expressing my concerns about the bill and their impact to Netizens and already marginalised communities including sex workers.

And while the current eSaftey Commissioner has stated that the sex industry is "not my concern" saying that “My role as a regulator is to protect all Australians from online harm — it’s not to restrict the sex industry. What happens between consenting adults is not my concern, as long as it’s not harming others, especially children,”[1] the expectations are **principles-based** and intended to be read in their broadest sense, which only enforces my concern that the eSaftey Commissioner has broad discretion to determine what is in the public interest.

[1] https://www.sbs.com.au/news/the-feed/rushed-through-parliament-sex-workers-concern-about-online-safety-act

Provider Will Consult with Commissioner and Refer to Commissioner's Guidance in Determining Reasonable Steps to Ensure Safe Use

Core expectation - In determining what are reasonable steps for the purposes of subsection 6(1), the provider of the service will consult the Commissioner.

One of the problems I see with the principals in their current state is that they are stated in a such a way that does not make it clear if an organization is meeting an Expectation and so the "service provider will be expected to consult with the eSafety Commissioner to determine what reasonable steps means for that provider".

While I commend that eSafety Commissioner and the eSafety Commissioner's office for providing the opportunity for service providers to consult with the government on these matters, it indicates that the eSafety Commissioner is aware that the principals and Expectations are not clear enough for providers to properly understand if they have met their obligations.

In the long term this can be problematic, as providers will be poorly situated to navigate subsequent decisions in a way that allows them to ensure they continue to meet the Expectations intent or navigate its challenges without close consultation with the eSafety Commissioner putting a strong dependence on this role, which as I outlined in my previous submission had very few transparency requirements.

As this feedback relates to a core Expectation, I understand that this cannot be amended as part of the determination, however, it would be excellent to see the underlying principals and Expectations revisited and redefined to clearly state the desired outcomes of adopting the principal or meeting the Expectation.

Furthermore, the eSafety Commissioner should publish specific outcomes that are desirable based on the classification of the provider, or services they provide -- this helps to ensure that controls that providers put in place reduce harm and do not just check a box for the Commissioner's office.

Provider Will Take Reasonable Steps to Ensure Safe Use

Core expectation -- The provider of the service will take reasonable steps to ensure that end-users are able to use the service in a safe manner.

Additional expectation -- The provider of the service will take reasonable steps to proactively minimise the extent to which material or activity on the service is or may be unlawful or harmful.

Furthermore, I am concerned that these vague principals and Expectations, which are intended to be read in the broadest sense, are then thrust upon service providers, which

once introduced are not enforceable by proceedings in court, but non-compliance with the Expectations could still potentially carry significant reputational consequences and civil penalties of up to AUD 111,000 for not complying with a notice or a determination relating to the reporting requirements.

This is likely to result in a similar affect that was seen when President Trump signed into law FOSTA, the Fight Online Sex Trafficking Act, and the Senate bill, SESTA, the Stop Enabling Sex Traffickers Act -- where providers and users of interactive computer service were suddenly treated as the publisher or speaker of any information provided by another information content provider -- meaning that website publishers would be held responsible for the misuse of services by third-parties[2].

Instead of making it easier to cut down on illegal sex trafficking online, the passing of these bills resulted in numerous providers taking blanket to censor or ban parts of their platform in response, not because their platforms were promoting sex trafficking, but because policing the platform while also facing significant consequences if they were found to be advertising or hosting material related to sex trafficking, meant it was a better investment to remove the service rather than trying to police it.

And while many large platforms and search engines go to great lengths to proactively detect and remove unlawful and harmful media, this approach (a monitoring and detection requirement) would not be appropriate or practical for all industry participants and would require further discussion.

The additional Expectation must better define what "unlawful" and "harmful" material is, if this additional expectation defines "unlawful" or "harmful" as anything that falls into the material defined in ***Expectation 11 - Provider will take reasonable steps to minimise provision of certain material***:

- (a) cyber-bullying material targeted at an Australian child;
- (b) cyber-abuse material targeted at an Australian adult;
- (c) a non-consensual intimate image of a person;
- (d) class 1 material;
- (e) material that promotes abhorrent violent conduct;
- (f) material that incites abhorrent violent conduct;
- (g) material that instructs in abhorrent violent conduct;
- (h) material that depicts abhorrent violent conduct.

Then this should be made clearer. If "unlawful" or "harmful" is defined differently, this must be clearly stated else service providers may begin to proactively remove material that is deemed to fall into these categories either by content moderators, provider employees, or their board members, effectively giving private organizations broad discretion to determine what is morally acceptable and beneficial, which we already

[2] https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom

know already disproportionately harms BIPOC (Black, Indigenous, People of Colour), LGBTQ+ folks, fat people, and sex workers.

For example, under the Expectations there is no indication as to whether spreading COVID-19 misinformation is considered harmful, and there is no clear definition of whether promoting racist-ideals, and white supremacy is considered harmful.

The most troubling part of this Determination is that the government claims that a key principle underlying the Act is that "the rules and protections we enjoy offline should also apply online" while also giving private organizations like Facebook who were found to be fuelling hate speech and violence in India[3] and Google who in 2019 was found to have created a hate speech detection bot that had a racial bias[4] the power to determine what is "unlawful" or "harmful".

Without clear guidance about what "unlawful" or "harmful" means, the inclusion of the additional expectation undermines the core expectation of the BOSE: The provider of the service will take reasonable steps to ensure that end-users are able to use the service in a safe manner.

Lastly, the recommendations made in Global Partners Digital & Digital Rights Watch join submission surrounding mitigating the risks associated with automated processes and artificial intelligence should be strongly considered as part of a robust harm reduction strategy.

Provider Will Take Reasonable Steps Regarding Encrypted Services

> *Additional expectation - If the service uses encryption, the provider of the service will take reasonable steps to develop and implement processes to detect and address material or activity on the service that is or may be unlawful or harmful.*

For more than 20 years, governments in Australia, the US, and the UK have been arguing against encryption and its advocates claiming that there is no legitimate reason for everyday public use of encryption. Even though it has transparently provided countless benefits to every person who has access to digital spaces from keeping your banking details private, to allowing domestic violence victims to securely coordinate their escape from abusive households.

The addition of this Expectation frames encryption as the adversary of online safety which runs counter to a research paper[5] authored by some of the best-known computer

[3] https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/
[4] https://fortune.com/2019/08/16/google-jigsaw-perspective-racial-bias/
[5] https://www.schneier.com/cryptography/paperfiles/paper-keys-under-doormats.pdf

security researchers which explored the problems with implementing such policies in practice.

In this paper they concluded "the damage that could be caused by law enforcement exceptional access requirements would be even greater today than it would have been 20 years ago." Such schemes kill innovation.

Further to this, encryption is used at various points in computer systems to help providers ensure personal information that is held is not subject to misuse, loss or to unauthorised access, modification, or disclosure. It is also an important part of harm reduction, as if data is breached or leaked the hard caused to an individual can include:

- Reputational damage
- Embarrassment or humiliation
- Emotional distress
- Identity theft or fraud
- Financial loss
- Loss of employment or business opportunities
- Family violence
- Other physical harm and intimidation
- Disruption of government services
- Unwanted marketing and spam email

With many of these we know disproportionately harm minorities such as women, BIPOC (Black, Indigenous, People of Colour) and LGBTQ+ folks.

As it is currently worded, this Expectation should be removed, because of the risks that is poses to digital security and because the government has failed to demonstrate the necessity of this obligation -- especially considering the accompanying FAQ document outlines that providers could meet this Expectation via behavioural, account or online signals including routing information and metadata. Without the clarification of the FAQ document which is not formally a part of the Determination it leaves room for the Expectation to be later interpreted or enforced to undermine encryption.

Instead, an Expectation requiring providers to make easily accessible and strong user-reporting systems, which encourage users to report abuse, publish and meet response-time service level agreements, which should be established with consultation, finally providers should be required, to publish and maintain an online safety policy which should include details about response times, and what users can expect it terms of redress.

Provider Will Take Reasonable Steps Regarding Anonymous Accounts

Additional expectation - If the service permits the use of anonymous accounts, the provider of the service will take reasonable steps to prevent

> *those accounts being used to deal with material, or for activity, that is or may be unlawful or harmful.*
>
> *Reasonable steps that could be taken - Without limiting subsection (1), reasonable steps for the purposes of that subsection could include the following:*
>
> - *(a) having processes that prevent the same person from repeatedly using anonymous accounts to post material, or to engage in activity, that is unlawful or harmful;*
> - *(b) having processes that require verification of identity or ownership of accounts.*

At present many popular websites such as Facebook and Twitter, require users to verify ownership of their account either via email or mobile number, however, it is generally well known that these processes can be easy to manipulate for users who wish to remain anonymous. Users who are often a part of vulnerable communities such as political activists, journalists, survivors of domestic abuse and LGBTQ+ folk.

Other methods of user identification also exist, however, often require individuals to provide identifying information or documents such as passports, drivers licenses or other key identification documents that are then cross-referenced or supplied to third-party identification systems. However, these third-party systems pose risks to individuals' privacy and personal security, including through data leaks or the mismanagement of personal information.

For example, the _Notifiable Data Breaches Report: January–June 2021_ published by The Office of the Australian Information Commissioner (OAIC) indicated[6] 446 breaches were notified under the scheme, with malicious or criminal attacks remaining the leading source, accounting for 65% of the total, while 30% of reports were the cause of human error.

The most concerning finding of this report when considered as part of this Expectation is that most data breaches (91%) notified under the NDB scheme involved 'contact information', such as an individual's name, home address, phone number or email address. While 'identity information', which was exposed in 55% of data breaches and includes an individual's date of birth, passport details and driver license details.

These statistics demonstrates that most providers are very unlikely to be able to handle contact and identity information in a way that doesn't compromise the security and safety of vulnerable communities such as political activists, journalists, survivors of domestic abuse and LGBTQ+ folk. And so, restrictions on anonymity, as proposed in this

[6] https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-statistics/notifiable-data-breaches-report-january-june-2021

Expectation do not clearly meet the core principle of ensuring that end-users are able to use the service in a safe manner.

Furthermore, this does not consider the impact identification requirements may have on those who may not have official forms of identification. Identification requirements also introduce additional risk to marginalized groups or those currently subjected to domestic violence as they very often will have their identification documents withheld from them or tightly controlled. As such, by enforcing identification requirements these groups of people are more likely to be monitored or face further abuse if discovered.

While many op-eds and opinion pieces have been published on the prevalence of harm perpetuated by anonymous accounts vs non-anonymous accounts, there is very little data on whether anonymity encourages or incites people to engage further in abuse online. However, research has found that anonymity is often a tool that empowers and protects individuals[7], furthermore, one of their recommendations included:

> The use of encryption and anonymity tools and better digital literacy should be encouraged. The Special Rapporteur, recognizing that the value of encryption and anonymity tools depends on their widespread adoption, encourages States, civil society organizations and corporations to engage in a campaign to bring encryption by design and default to users around the world and, where necessary, to ensure that users at risk be provided the tools to exercise their right to freedom of opinion and expression securely.

Therefore, the additional expectation and reasonable steps be removed from the Act. Instead, expectations and reasonable steps should focus on encouraging platforms to provide more power to users to select who they do and do not interact with -- for example, allowing users to restrict their interaction with anonymous accounts -- rather than forcing users to verify their identity using methods that may undermine their privacy, and personal safety.

Provider Will Keep Records Regarding Certain Matters

> *The provider of the service will keep records of reports and complaints about the material mentioned in section 13 provided on the service for 5 years after the making of the report or complaint to which the record relates.*

As mentioned as a concern in the previous section about identity verification, the ***Notifiable Data Breaches Report: January–June 2021*** published by The Office of the Australian Information Commissioner (OAIC) indicated[8] 446 breaches were notified

[7] https://www.undocs.org/A/HRC/29/32

[8] https://www.oaic.gov.au/privacy/notifiable-data-breaches/notifiable-data-breaches-statistics/notifiable-data-breaches-report-january-june-2021

under the scheme, with malicious or criminal attacks remaining the leading source, accounting for 65% of the total, while 30% of reports were the cause of human error.

The most concerning finding of this report when considered as part of this Expectation is that most data breaches (91%) notified under the NDB scheme involved 'contact information', such as an individual's name, home address, phone number or email address. While 'identity information', which was exposed in 55% of data breaches and includes an individual's date of birth, passport details and driver license details.

While I do not believe it is unreasonable that providers should keep records of reports and complaints, the eSaftey Commissioner must ensure that providers are given appropriate guidance and support to ensure that this data is stored securely and kept private to ensure the safety of the complainant as well as the person impacted by the complaint, whether content was removed or not.

The eSaftey Commissioner may consider adding in reasonable steps that see that providers introduce data protection standards such as providing clear advice on the minimum set of data that must be retained by businesses, define a minimum set of protection standards such as requiring encryption at rest and requiring restricted access to saved reports, and defining how to dispose of data at the end of the defined five-year period.