Director, Online Safety Reform and Research Section
Department of Infrastructure, Transport, Regional Development and Communications
GPO Box 2154
Canberra ACT 2601
OnlineSafety@infrastructure.gov.au

Tuesday 31 August 2021

re: Draft Online Safety (Basic Online Safety Expectations) Determination 2021 consultation

Dear Director,

I am writing in response to the exposure draft of the Online Safety (Basic Online Safety Expectations) Determination 2021.

Yours faithfully,
[REDACTED]

Preface: Online Safety Act 2021

The Online Safety Act 2021 ("the Act") purports to keep Australians safe online, but is a destructive and paternalistic piece of legislation that poses a serious threat to the internet as well as Australians' freedom of expression, access to information, privacy, security and democracy.

> "A key principle underlying the Act is that the rules and protections we enjoy offline should also apply online." *(Draft Online Safety (Basic Online Safety Expectations) Determination 2021 consultation)*

This premise of the Act is flawed for multiple reasons.

- It downplays the fact that the offline and online worlds are fundamentally different.
- It seems to imply that online lacks regulation, despite many existing laws that apply online.
- Many rules and protections introduced by the Act go far beyond those that apply offline.

Online spaces can be misused to cause direct harm to people, such as fraud, doxxing and psychological harm; and to coordinate, facilitate or induce offline dangers, such as violence and abuse. However, beyond that, the offline–online safety parity argument falls apart. Physical forces do not transmit through the internet, while information can be easily copied and tends to travel much faster and wider online. Furthermore, people may have online presence but fundamentally exist offline.

The government and service providers are not our parents. Governments have repeatedly abused their powers[1,2], and social media providers have demonstrated they are incapable of moderating content[3,4]. Despite that, the Act effectively puts the government (mainly the eSafety Commissioner) and service providers (by compulsion) in parenting roles over all Australians' use of online spaces. The Act also claims extraterritorial jurisdiction to impose its nanny-state measures onto service providers globally.

1 Ben Smee. "Queensland police officer receives suspended jail sentence for leaking woman's details to violent ex-husband". October 2019. https://www.theguardian.com/australia-news/2019/oct/14/queensland-police-officer-pleads-guilty-leaking-womans-details-violent-ex-husband

2 Bernard Keane. "Data retention scheme is being abused exactly as critics predicted". February 2020. https://www.crikey.com.au/2020/02/25/data-retention-scheme-abuse/

3 Lacey-Jade Christie. "Instagram censored one of these photos but not the other. We must ask why". October 2020. https://www.theguardian.com/technology/2020/oct/20/instagram-censored-one-of-these-photos-but-not-the-other-we-must-ask-why

4 Onlinecensorship.org. "Offline-Online". https://onlinecensorship.org/content/infographics

The Act's approach is not only techno-solutionist[5,6] and extremely paternalistic, it also merely treats the symptoms of social problems but not their root causes. The best approach to harm reduction of online spaces without introducing other harms is debatable, but it is Australian communities and families who must assume primary responsibility to tackle social challenges and promote safe and disciplined use of technology.

Although the Act is capable of improving online safety in some ways, it also would put Australians at risk of other harms. The Act, with the proposed Basic Online Safety Expectations, would:

- give the eSafety Commissioner and its delegates sweeping powers void of accountability and transparency, including immunity from liability,
- incentivise or effectively require (depending on level of enforcement) that services proactively remove material using automated tools[7],
- prohibit access to various material, notably "class 1" material, with insufficient exemptions,
- effectively require mechanisms such as facial recognition[8] to gate-keep "class 2" materials,
- systemically undermine end-to-end encrypted communications, and
- require that anonymous account users identify themselves or else face restrictions;

which would apply to almost every digital communication that touches Australia.

These measures risk introducing or worsening material and psychological harms such as rampant censorship, chilling effects, privacy violations, normalisation of surveillance and extreme security risks. The Act risks making the internet a hostile and potentially dangerous place for Australians, especially for marginalised and vulnerable groups of people[9]. Furthermore, the sweeping and unaccountable censorship powers would undermine Australians' freedom of expression and access to information, which are vital for democracy.

It appears that the Act's delegated legislation is being used tactically as a legislative backdoor to introduce unpopular measures the government wants without parliamentary or citizen scrutiny. A notable example is backdoors into end-to-end encryption, which Five Eyes governments are fiercely pushing for[10,11] despite persistent opposition by Australians[12] and civil society worldwide[13]. The Act and its associated bills and explanatory memoranda make no mention of "encryption", but after the Act was solidified into law, "encryption" appears in the exposure draft of the Basic Online Safety Expectations. This raises serious concerns that the Act is drafted and intended to introduce backdoors into end-to-end encryption *by design*.

---

5 Ashali Bhandari. "Feminist Perspectives on Space, Safety and Surveillance: Improving a Woman's Right to the City". March 2021. https://thewire.in/women/feminist-perspectives-on-space-safety-and-surveillance-improving-a-womans-right-to-the-city

6 Digital Rights Watch. "Policy grounded in surveillance won't protect women". June 2021. https://digitalrightswatch.org.au/2021/06/16/policy-grounded-in-surveillance-wont-protect-women/

7 Electronic Frontier Foundation. "Facebook's Most Recent Transparency Report Demonstrates the Pitfalls of Automated Content Moderation". October 2020. https://www.eff.org/deeplinks/2020/10/facebooks-most-recent-transparency-report-demonstrates-pitfalls-automated-content

8 Ariel Bogle. "Porn age filter for Australia recommended by parliamentary committee". March 2020. https://www.abc.net.au/news/science/2020-03-05/age-verification-filter-for-online-porn-recommended-in-australia/12028870

9 Grace O'Brien. "Racial Profiling, Surveillance and Over-Policing: The Over-Incarceration of Young First Nations Males in Australia". Published February 2021. https://www.mdpi.com/2076-0760/10/2/68

10 Department of Justice (US). "International Statement: End-To-End Encryption and Public Safety". October 2020. https://www.justice.gov/opa/pr/international-statement-end-end-encryption-and-public-safety

11 Julie Inman Grant. "End-to-end encryption: a challenging quest for balance". February 2020. https://www.esafety.gov.au/about-us/blog/end-end-encryption-challenging-quest-for-balance

12 Digital Rights Watch. "Research shows Australians deeply concerned by Government's spyware legislation". October 2018. https://digitalrightswatch.org.au/2018/10/25/research-shows-australians-deeply-concerned-by-governments-spyware-legislation/

13 "Security for All". https://securetheinternet.org/

The Act takes a disproportionately punitive approach towards end-users and service providers. Alarmingly, service providers and end-users could be fined 500 penalty units (currently $111,000) and potentially face other punishments in relation to several parts of the Act, and service providers must be able to comply with notices within 24 hours to avoid the civil penalty. Appropriate guidance or warnings, not punishments, should be given to end-users and service providers who put effort into corrective action or compliance with the Act. Punishments should be proportionate and reserved for negligence, refusal to comply and repeat offences.

Although the Act provides a few commendable measures for online safety, I strongly believe the Act must be repealed in its entirety. The government must engage with Australians in a genuine and meaningful dialogue about how it can play a role to reduce online harms, while preserving Australia's human rights, civil liberties, democracy and security, not simply ram the laws it wants through parliament. I make my submission in relation to the Basic Online Safety Expectations while maintaining my position of opposition to the Act.

Basic Online Safety Expectations

Note: In the exposure draft, the ordering and numbering of sections under contents and in the body differ. My submission assumes the numbered sections in the body are what should be referred to.

The Basic Online Safety Expectations ("BOSE") imposes expectations on social media services, relevant electronic services and designated internet services (collectively "service providers").

Expectations: ensure safe use (section 6)

By requiring *proactive* measures against material or activity that *may be* unlawful or harmful, the additional expectation (2) and suggested reasonable step (3)(a) would establish an aggressive censorship bias, and incentivise or effectively require that service providers use automated tools.

The combination of automated tools and service provider moderation mistakes (personal biases and lack of community or cultural understanding) poses an unacceptable risk that innocuous materials will be wrongly taken down. This is especially concerning given the Act generally requires rapid take-down of material or activity but not rapid and easy put-back upon false positives.

Recommendation 1: Remove "proactively" and "or may be" from additional expectation (2).

Recommendation 2: Avoid incentivising or effectively requiring the use of automated tools to identify and remove material or activity.

Recommendation 3: Require reasonable steps to segregate class 2 material and exempt material that would otherwise be subject to section 11 from all other material, such that access to segregated material is strictly opt-in and comes with clear and appropriate content warnings. Segregated material may optionally be access-controlled by parental control features, or further segregated by category (such as sex, gambling, drugs, violence, etc.) so that fine-grained opt-in is possible.

Recommendation 4: Add to steps in (3): encouraging privacy, safety and security best practices.

Recommendation 5: Replace reasonable step (3)(b) with: ensuring that the default privacy, safety and security settings of the service are robust and set to the highest possible level.

Additional expectation: encrypted services (section 8)

The way this expectation is worded, and given the background and context of this expectation, it appears that "encryption" likely refers to end-to-end encryption (E2EE).

This expectation effectively requires that backdoors be built into E2EE.

Five Eyes governments[14] and the Commissioner[15] are on a crusade against E2EE in the guise of striving for "balance". The Act and its associated bills and explanatory memoranda make no mention of "encryption", but after the Act was solidified into law, "encryption" appears in this draft of delegated legislation in the form of this expectation. This raises serious concerns that the Act is drafted and intended to introduce backdoors into E2EE *by design*.

There is no such thing as "balance" in relation to E2EE; the technology is either secure against third-party access, or it is not[16]. E2EE and this expectation are mutually contradictory.

Development of detecting offending images (by perceptive hash matching) that preserves security of E2EE was attempted[17], but after its development the authors warned their detection scheme undermines E2EE[18]. The biggest problem with the detection scheme is that the matching database can easily be swapped from a list of genuine offending images (such as child exploitation material and extremist propaganda) to another list of images (such as dissident material, police brutality footage and evidence of wrongdoing) in order to identify dissidents, journalists, whistleblowers and human rights defenders. Text-based schemes would suffer similar vulnerability. Actors who could exploit the scheme would not be limited to the service provider or government that controls the matching database, but extend to malicious end-users and unauthorised actors outside the scheme.

E2EE is a necessary technology that affords security and privacy to everyday people, and is especially vital for a wide variety of high-risk groups such as dissidents, journalists, whistleblowers, human rights defenders, activists, people with personal safety fears, marginalised groups, business executives, government officials, doctors and lawyers. Without E2EE that is secure, high-risk people are chilled into silence[19] or risk putting their safety at risk if they speak out, and confidentiality of business and privileged communications risk being breached.

The recent revelation about spyware "Pegasus" demonstrates that the surveillance technology industry develops spyware in the guise of "fighting crime" and sells them to questionable governments with poor human rights records, who then use the spyware to go after dissidents, journalists and human rights defenders[20]. Hacking into devices poses a serious threat to targetted people and causes a chilling effect on society as a whole. Similarly, side-stepping E2EE by design to gain third-party access is dangerous.

This expectation appearing in the final version of BOSE would set a dangerous precedent for further attacks against E2EE in Australia and worldwide. This absolutely must not be allowed.

Recommendation 6: Remove this expectation.

Recommendation 7: Require that no mechanism is implemented that in any way weakens, side-steps, ghosts[21], key escrows, interferes with or otherwise undermines end-to-end encryption or any other security or privacy feature related to end-to-end encrypted communication.

---

14 Department of Justice (US). "International Statement: End-To-End Encryption and Public Safety". October 2020. https://www.justice.gov/opa/pr/international-statement-end-end-encryption-and-public-safety

15 Julie Inman Grant. "End-to-end encryption: a challenging quest for balance". February 2020. https://www.esafety.gov.au/about-us/blog/end-end-encryption-challenging-quest-for-balance

16 Harold Abelson et al. "Keys Under Doormats: Mandating insecurity by requiring government access to all data and communications". July 2015. https://dspace.mit.edu/bitstream/handle/1721.1/97690/MIT-CSAIL-TR-2015-026.pdf

17 Anunay Kulshrestha and Jonathan Mayer. "Identifying Harmful Media in End-to-End Encrypted Communication: Efficient Private Membership Computation". 2021. https://www.usenix.org/conference/usenixsecurity21/presentation/kulshrestha

18 Anunay Kulshrestha and Jonathan Mayer. "We built a system like Apple's to flag child sexual abuse material — and concluded the tech was dangerous". August 2021. https://www.washingtonpost.com/opinions/2021/08/19/apple-csam-abuse-encryption-security-privacy-dangerous/

19 Jillian York. "The harms of surveillance to privacy, expression and association". 2014. https://giswatch.org/thematic-report/internet-rights/harms-surveillance-privacy-expression-and-association

20 Forbidden Stories. "About The Pegasus Project". July 2021. https://forbiddenstories.org/about-the-pegasus-project/

21 Ghosting means inserting an endpoint into a communication without the corresponding parties' control or knowledge.

Additional expectation: anonymous accounts (section 9)

The Act and BOSE do not define "anonymous account". Given reasonable step (2)(b) is incompatible with unverified accounts, it appears "anonymous account" means an account with no profile or whose profile does not resemble the end-user's identity ("non-realname account").

This expectation requires that non-realname accounts face increased scrutiny or restrictions. Given that the Act requires minimisation of material or activity that is or may be unlawful or harmful regardless of what accounts are used, there is no logical reason to apply increased scrutiny or restrictions onto non-realname accounts.

Reasonable step (2)(b) suggests requiring identity verification of non-realname accounts. If this suggestion is enforced sufficiently, many Australians would need to submit biometrics or ID in order to start or continue using social media and other online services.

This identity verification policy is neither necessary nor proportionate. In the offline world, Australians generally do not need to show ID when they go shopping, play sport, attend events, go to church, see friends or perform most other daily activities. The identity verification policy risks making the online world a "show me your papers please" society for Australians.

Furthermore, reasonable step (2)(b) appears to incentivise or effectively require an automated mechanism such as biometric recognition or identity document scanning. This may involve service providers collecting biometric or identity information, or use of the government's centralised biometric recognition and identity database "The Capability".

These mechanisms pose extreme security risk for end-users. Maintaining databases of biometric and identity information increases the security risks of data breach[22] such as accidental disclosure, rogue insider and malicious hacking, severely changing the lives of people affected.

Many of the points raised in the previous section about the importance of E2EE also apply to anonymity. Anonymity allows people to express their genuine opinions and political beliefs to the world without fear of retribution; this is why Australian elections use the secret ballot. Anonymity allows people to go about their lives in public space without their activities being tracked and analysed. Anonymity allows dissidents, journalists, whistleblowers, human rights defenders, business executives and marginalised groups to exist without being targetted by adversarial actors.

Considering the security and privacy risks of identity verification, as well as the benefits of anonymity, identity verification for this expectation's intended purpose is clearly unacceptable.

Recommendation 8: Remove this expectation.

Recommendation 9: Require that no mechanism is implemented that in any way identifies, fingerprints, analyses identifying characteristics of or otherwise undermines the anonymity of an anonymous or pseudonymous end-user or activity; unless an advertised service feature requires such mechanism, in which case affected anonymous or pseudonymous end-users must voluntarily, affirmatively, informedly and specifically consent to use of the mechanism.

Core expectation: minimising provision of certain material (section 11)

Although the Act (section 104) specifies exemptions that apply to material related to abhorrent violent conduct, the exemptions are missing in the exposure draft of BOSE.

The exemptions cover only abhorrent violent conduct, and those exemptions don't go far enough. The public interest exemption is narrow in scope, applying only to materials made by a professional journalist in relation to a news report or current affairs story, therefore would not exempt

22 Kevin Nguyen. "NSW driver's licence data breach left Sydney health worker 'sickened'". September 2020. https://www.abc.net.au/news/2020-09-02/sydney-man-finds-own-driver-licence-in-nsw-data-breach/12616606

recordings made by witnesses of incidents uploaded in the public interest. Sex work, a lawful occupation in Australia, is notably absent from being exempted, therefore a high-risk target of being censored.

Many service providers are ill placed to determine what material is bullying, abusive, consensual, abhorrently violent, journalistic, for research, educational or genuinely artistic. Furthermore, the Act pressures service providers to over-censor material or activity in order to remain compliant, including by use of automated tools.

The Act relies too heavily on Australia's classification system for categorisation of class 1 (RC) and class 2 (X18+ and R18+) material. Australia's classification system has a reputation of being one of the strictest in the Western world (especially towards video games), and dictates what material Australians (including adults) can and cannot access. Furthermore, the Act gives the Commissioner broad discretion to determine whether each unclassified material would fall under class 1 or class 2 or otherwise.

The combination of the issues described above would inevitably lead to unacceptably high rates of innocuous material being wrongly taken down by service providers.

Recommendation 10: Add the exemptions specified under the Act (section 104) to this expectation.

Recommendation 11: Add exemptions to all material that is in the public interest (by a professional journalist or otherwise), sex work, artistic, for research, educational or for advocacy.

Core expectation: preventing access by children to class 2 material (section 12)

This expectation, with suggested reasonable step (2)(a), appears to incentivise or effectively require an automated mechanism such as biometric recognition or identity document scanning. Using facial recognition technology for this purpose[23] was suggested by the department in charge of a centralised database "The Capability" designed to catalogue biometric, identity and other information on every Australian.

Technology involving biometric recognition or identity document scanning is highly intrusive and poses an extreme security risk for end-users. Furthermore, combining this intrusive technology with The Capability would make the government capable of tracking the activities of every Australian who engages with the system. This is completely unacceptable.

Recommendation 12: Prohibit the use of biometric recognition tools, identity document scanning tools or similar mechanisms that pose privacy or security risks to end-users. Perhaps the safest (but least practical) method is offline ID citation, but any secure method that the end-user can use to prove they are not a "child" without collecting any other information about them is acceptable.

Additional expectation: provide regular reminders about policies (section 18)

This expectation requires that service providers ensure that end-users receive regular reminders about the service's policies. This would likely lead to unpleasant user experiences for end-users.

End-users should be able to choose whether or not they receive reminders. The expectation should also be reworded such that sending is ensured, not receiving. I support the end-user being notified of non-trivial changes to service policies with some minimum notice period.

Recommendation 13: Allow end-users to choose whether or not they receive regular reminders.

Recommendation 14: Require the service provider to ensure they send reminders, not ensure that end-users receive them.

23 Chris Duckett. "Home Affairs pushes its face-matching service for porn age verification". October 2019. https://www.zdnet.com/article/home-affairs-pushes-its-face-matching-service-for-porn-age-verification/

Recommendation 15: Require the service provider to send a notification to end-users of any non-trivial change to the service's policy with some minimum notice period.

Additional additional expectations

Recommendation 16: Require the existence of an appeals mechanism that is accessible to end-users and allows for rapid restoration of material or activity in case of false positives.

Recommendation 17: Require the restoration of material or activity without undue delay when its removal is successfully appealed or when served with a notice revocation.

Recommendation 18: Require that all systems used for the purpose of the Act (such as BOSE section 12) use transparent standards, use free and open-source software instead of proprietary software, and be subject to independent oversight. Additionally, all software distributed over a network for immediate use (such as on an end-user's device) must be copyleft licensed such that the network-used software's freedoms are protected[24].

Expectations for the Commissioner

The following expectations for the Commissioner are proposed. These recommendations also apply to the Commissioner's delegates.

Recommendation 19: Require the Commissioner to act with due care, diligence and skill in relation to all its functions and powers under the Act.

Recommendation 20: Require the Commissioner to establish an appeals mechanism that does not involve the court system, is accessible to end-users and service providers, and allows for rapid restoration of material or activity in case of false positives.

Recommendation 21: Require the Commissioner to issue a revocation of a successfully appealed notice within 24 hours of the appeal.

Recommendation 22: Require the Commissioner to regularly report to the public in relation to all notices issued under the Act. This reporting must include numbers of notices issued, revoked and appealed; in total, by each type, by each recipient ("end-users" as one statistic), and by each category of material or activity.

Recommendation 23: Require the Commissioner to regularly update a public list of every item (material, software ("apps"), links, domain names and IP addresses) affected by a notice under the Act. This list must include for each item: issuance/revocation date, notice type, whether the notice was issued or revoked, reason for issuance/revocation, an item description that allows identification of the item, and on which service the item existed prior to the notice.

Recommendation 24: Require the Commissioner be subject to regular oversight by a parliamentary oversight board in relation to all activity covered by the Act.

Recommendation 25: Require the Commissioner be subject to regular oversight by a multi-stakeholder community oversight board in relation to all activity covered by the Act.

Recommendation 26: Require the Commissioner to apply to the Australian Classification Board for classification of an unclassified material if the Commissioner determines that the material would be classified in a way that makes the material fall under class 1 or class 2.

24 Choose a License. https://choosealicense.com/appendix/#network-use-disclose