The Cabinet Office

NSW Government Submission to the Statutory Review of the Online Safety Act 2021 (Cth)

June 2024

NSW GOVERNMENT

Acknowledgement of Country

The Cabinet Office acknowledges the Traditional Custodians of the lands where we work and live. We celebrate the diversity of Aboriginal peoples and their ongoing cultures and connections to the lands and waters of NSW.

We pay our respects to Elders past, present and emerging and acknowledge the Aboriginal and Torres Strait Islander people that contributed to the development of this submission.

We advise this resource may contain images, or names of deceased persons in photographs or historical content.

Acknowledgements

This submission has been prepared by The Cabinet Office (**TCO**) with contributions from the NSW Police Force (**NSWPF**), the NSW Ministry of Health (**Health**), the NSW Department of Education (**NSW DoE**), the NSW Department of Communities and Justice (**DCJ**), the NSW Department of Customer Service (**DCS**), and Multicultural NSW (**MNSW**).

Contents

1 Introduction

The NSW Government welcomes the Australian Government's review of the *Online Safety Act 2021* (Cth) (**OSA**) (**the Review**).

Online platforms are a core part of many Australians' daily lives. Online environments provide a range of benefits through increased social connection, civic engagement opportunities, educational resources, access to information and entertainment, online tools, and the digital economy. It is imperative that these digital spaces are safe and inclusive, to protect vulnerable populations, foster social cohesion, and support mental health, public safety and security.

The NSW community has recently experienced the impacts of online harms in the wake of the critical incident at Bondi Junction on 13 April 2024 and the declared terrorist incident at a Wakeley church on 15 April 2024.

Following the Bondi Junction incident, distressing photos and videos were widely disseminated on social media, along with speculation about the perpetrator's motives. Following the livestreamed attack on members of the clergy at the Wakeley church, a major public disorder incident occurred outside the church. Social media posts generated in relation to the public disorder incident have had a detrimental impact on community cohesion.

The NSW Government is committed to partnering with the Australian Government and participating in national approaches to ensure online safety. We acknowledge the breadth of work the Australian Government is delivering alongside this review to protect against online harms, including reviewing the National Classification Scheme and developing mandatory industry codes in relation to scam activity.

Similarly, the NSW Government is also progressing work to protect the community from online harms, including hosting a Social Media Summit this year. The Summit will provide an opportunity to generate innovative solutions to the challenges of social media and inform potential regulatory and legislative changes.

Other policies the NSW Government has put into place to minimise the negative impacts of social media and devices on young people include:

- a mobile phone ban in all NSW public schools implemented in October 2023
- a $2.5 million research fund to investigate the impacts of excessive screen time, video games and mobile phone use on young people and their learning
- recruiting 250 additional school counsellors
- a review into evidenced-based practice and school policy which can address school student's online behaviour, led by NSW Chief Behaviour Advisor Professor Donna Cross.

The NSW Government supports consideration of limiting access for children and young people to social media and would welcome discussions with the Australian Government on piloting age verification technology. NSW has successfully trialled age verification for the online purchase of alcohol, in partnership with Tipple and Mastercard, and could provide learnings from this experience.

The Parliament of NSW's Joint Standing Committee on Electoral Matters is conducting an *Inquiry into the Administration of the 2023 NSW State Election and Other Matters*, which is considering mis-and dis-information in an electoral context. This inquiry is due to report in mid-2025. The NSW Government has also asked the Law Reform Commission to review the effectiveness of section 93Z of the *Crimes Act 1900* (NSW) in addressing serious racial and religious vilification.

Overview of the submission

The Cabinet Office (**TCO**) has prepared this submission with contributions from a range of NSW agencies including the NSW Police Force (**NSWPF**), the NSW Ministry of Health (**Health**), the NSW Department of Education (**NSW DoE**), the NSW Department of Communities and Justice (**DCJ**), the NSW Department of Customer Service (**DCS**) and Multicultural NSW (**MNSW**).

The submission provides suggestions regarding the operation of the OSA in NSW, and responds to the consultation questions the Review has asked in the Issues Paper.

Key themes NSW has suggested the Review may wish to consider include:

- *the OSA's content-based settings*: these settings provide important recourse for those who have experienced online harms, however the settings for systems-based mitigation of online harms could be strengthened:
    - to better balance the responsibility for action between individual users, regulators and service/platform operators
    - for more efficient intervention to manage the scale of risk of harm upstream at the point of platform and algorithmic design rather than downstream on a content instance
    - for greater long-term regulatory effectiveness and resilience to rapid changes in technology and online activity
- *exploration of means to ensure compliance with existing laws for content hosted by companies located offshore*
- *eSafety Commissioner's powers and role:*
    - consideration of how the eSafety Commissioner's powers and functions could be used to investigate complaints about hate speech and mis/dis-information in online content increased communication to the public the role of the eSafety Commissioner and the support processes and resources available, particularly to those groups at greatest risk of victimisation.

2 Operation of the OSA in NSW

Since its commencement, the OSA has strengthened the ability of the NSWPF to request removal of harmful content by online service providers. It has also helped the NSW DoE support our schools in reporting serious online abuse on social media. The provisions enabling department staff to act on behalf of children to report image-based abuse and child cyberbullying have had a positive impact and work well.

Prior to the commencement of the OSA, the NSWPF made requests for the removal of content by online service providers and successful removal was determined by the individual provider's terms and conditions. Since the OSA's introduction, the NSWPF has had greater scope to request the removal of content from providers on the basis of its regulation of categories of harm.

The two most significant issues the NSWPF has had with the implementation of the OSA are:

1. when a contact is unavailable at the online service provider to assist with a content removal request
2. not receiving a response from the online service provider. There have been seven occasions where the NSWPF submitted requests to the provider and not received a response. As such the NSWPF is unable to determine what action the provider took in relation to these requests.

The establishment of the eSafety Commissioner provides the NSWPF with an additional pathway to apply to online service providers for the removal of content. However, it should be noted that this pathway only enables the eSafety Commissioner to request the same contact (within the provider) to remove content. In practice, this means that where a provider would have already responded to a NSWPF request by refusing removal because they do not believe it falls within the categories stipulated in the OSA, they will provide this same response to eSafety Commissioner. The eSafety Commissioner can take legal action against the online service provider, however this would not assist in urgently removing harmful material from circulation.

To address these issues, the NSWPF suggests that online service providers be required to:

1. have designated contacts that are always available to assist law enforcement, and when a contact is not available, an alternative contact person is provided. Alternatively, a generic email address that is always monitored by the online service provider could be provided
2. provide a response to law enforcement for every request made.

In February 2023, the NSWPF and eSafety Commissioner entered a memorandum of understanding (**MoU**) to formalise governance arrangements and protocols related to responsibilities under the OSA. The NSWPF has subsequently developed the eSafety Business Rules & standard operating procedures, to ensure the NSWPF meets its responsibilities under the OSA and the MoU.

Since this time there have been several occasions where the NSWPF has requested online service providers to remove content from their platforms, which is anticipated to increase. In this time, when NSWPF requested a provider remove content from their platforms under the OSA:

- the provider removed the content upon request by the NSWPF on
    - 12 occasions: without the assistance of the eSafety Commissioner
    - two occasions: with the assistance of the eSafety Commissioner
- the provider did not remove the content upon request by the NSWPF on

    - one occasion: NSWPF did not seek the assistance of the eSafety Commissioner
    - one occasion: NSWPF sought the assistance of the eSafety Commissioner
- The NSWPF was unsure whether the provider removed the content upon request by the NSWPF:
    - four occasions: NSWPF did not seek the assistance of the eSafety Commissioner
    - three occasions: NSWPF sought the assistance of the eSafety Commissioner.

NSW suggests that the following changes could improve implementation of the OSA in NSW:

- allowing for early notification to law enforcement agencies of the eSafety Commissioner's intent to remove content or issuing of notices under the OSA, in case these actions could prejudice ongoing investigations
- providing an escalation pathway when there are disputes over whether to release information to law enforcement when there is an imminent threat to life. These requests are made to social media services, relevant electronic services, designated internet services, internet search engine services, app distribution services and/or hosting services
- explicitly recognising criminal activity as harmful content, similar to what has been implemented in NSW 'post and boast' laws.

3 NSW Responses to Consultation Questions

3.1 Australia's regulatory approach to online services, systems and processes

Question 1. Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?

The current objects of the Act to improve and promote online safety for Australians is sufficient. However, there are gaps in regulated content that, if addressed, would help better achieve the objects of the Act.

The scope of regulated content could be expanded to provide a mechanism for people to lodge complaints for hate speech, mis/dis-information, and content that presents a public safety issue.

Ideally, the onus of responsibility for reducing spread of harmful content should lie with the service provider, rather than an individual. NSW supports the Review exploring a duty of care on service providers to online users to prevent the proliferation of harmful content, conduct and contact. Beyond source material, this would include re-sharing, comments and other user behaviours that amplify harmful material. This could extend to service provider recommender systems, moderation and advertising schemes. An enforceable Basic Online Service Expectations (BOSE) may be one method for doing so.

NSW suggests that the Review also consider whether the OSA should address:

- content relating to incitement to violence/public disorder that would present a public order issue other than lawful protest and public assembly
- content relating to misinformation campaigns that would present a public safety issue
- content that would facilitate the commission of serious crime
- content that would be deemed as doxing and would place a member of the public in danger or cause harm
- content or posts that facilitate access to capability documents and or manuals relating to the manufacture or explosives, firearms, prohibited drugs, law enforcement methodology, etc
- content or posts that would disclose the identity of Australian officials, which would otherwise be illegal to disclose such as the identification of Australian Security Intelligence Organisation (ASIO)/Australian Secret Intelligence Service (ASIS) officers under their respective Acts
- content or posts that are subject to non-publication/suppression orders made by a Court in Australia.

NSW also considers online content boasting about crimes to be harmful, particularly to young Australians. However, given the recent amendments to the *Crimes Act 1900* (NSW) to create an

offence to address this, there has not been sufficient time to determine whether further legislative change is required, including to the OSA (see response to Question 13 for further details).

The Review may also want to consider whether it could regulate content that does not meet the level of criminal behaviour, that nonetheless has negative impacts on young people and school communities such as student fight videos and vandalism.

Hate speech and mis/dis-information

The role of social media in amplifying narratives and conversations as part of the 'attention economy' is critical to a deficit in trust of government institutions, social cohesion, and optimism that are so importance for healthy democracies.

Localised online hate speech tends to increase during adverse global political events – and often exacerbates real-world tensions as a result.[1] It is important to consider that these real-world tensions can metastasize into real-world violence and can manifest across demographic settings: from school playgrounds to protest and social justice movements.

In a scan of social media in the first 24 hours following the declared terrorist incident at a Wakeley church on 15 April 2024, individuals online used the incident to provide a focal point to express and validate prejudiced beliefs, amplify divisive rhetoric and airing of racist sentiments, fuelling animosity among different social groups. Rapid dissemination of misinformation, speculation and conspiracy theories deepened existing grievances. Online commentary also intertwined issues of crime, mental illness, religious intolerance, driving global conflict narratives, spreading conspiracy theories, and exacerbating polarisation. There was anxiety over violence escalating in Sydney. However, positive messages, particularly from community leaders, can have a pivotal role in online conversations and for fostering unity.

Children are particularly vulnerable to the flow on effects of online hate speech induced by global trigger events. Recent research has revealed that young people who are exposed to online hate speech commonly experience negative feelings and may also be associated with processes of political radicalisation. Victims may lack appropriate coping strategies to mitigate the harm of hate speech and even seek revenge. As a current example in the NSW context, tensions arising from the ongoing Hamas-Israel war have impacted social cohesion in NSW public schools. The NSW DoE recently reported that sequential crises, including COVID-19, floods, fires, the Voice referendum, and the Hamas-Israel conflict, have significantly impacted student and teacher wellbeing in NSW.

The NSW DoE also reports that children are being exposed to highly traumatising and emotive material online (mostly through social media) because of the conflict — which is acutely impacting student wellbeing and increasing broader community polarisation.

While NSW understands the Australian Government is developing new legislation to provide the Australian Communications and Media Authority (**ACMA**) with powers to combat online mis/dis-information, we suggest consideration of a single complaints-based scheme to manage internet harms (including hate speech). The Review may want to take into consideration the role that hate speech can play in a continuum of (online and offline) violence (see the Institute for Strategic Dialogue's paper on Misogynistic Pathways to Radicalisation).

Given the global operation and reach of major online services and platforms, there may be benefits to adopting or adapting the World Economic Forum paper on Typology of Online Harms, developed by the Global Coalition for Digital Safety. This typology is being developed in recognition of the complexity of identifying highly local or context-specific harms, while balancing human rights, ethical, privacy, legal, social and technological considerations. It includes:

[1] https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

- threats to personal and community safety (e.g., child exploitation and sexual abuse, extremist and pro-terror including recruitment, violent graphic content – including incitement to violence, technology facilitated abuse and gender-based violence)
- harm to health (e.g. content promoting suicide and disordered eating)
- hate and discrimination
- violations of dignity and privacy (bullying, harassment, doxing and image-based abuse)
- deception and manipulation (e.g. mis and dis-information, impersonation, catfishing, deceptive synthetic media).

It also recognises the breadth of harms encompassed in the production of content (e.g. images of murder), the distribution of content (e.g. hateful comments towards certain populations), and the consumption of content (e.g. age-inappropriate content).

Question 4. Should the Act have strengthened and enforceable Basic Online Safety Expectations?

Unacceptable content and behaviour should be addressed at the source, by design, by those best placed to do so (in terms of capability, resources and reach). Intervention and sanctions by government should apply rapidly and decisively to any remaining unacceptable content.

A preventative approach first and foremost should be underpinned by an effective deterrence model, to avoid downstream practical enforcement difficulties.

The current system of the eSafety Commissioner needing to individually investigate and enforce the OSA is time-consuming and relies on individuals being aware of the Scheme and reporting. NSW suggests that more onus needs to be put on larger sites which operate under a free content model. This is particularly relevant for social media platforms that have a high number of young users and can potentially push users towards more pornographic or violent content.

In addition to current core expectations, NSW would suggest the Review consider expectations that cover:

- artificial intelligence (AI) and generative AI (GenAI) capabilities that are designed and implemented with user safety in mind, and that services using AI/GenAI capabilities proactively minimise the extent to which that capability produces unlawful or harmful material
- recommender systems that are designed and implemented in a manner that enables their safe use, and that services minimise the extent to which recommender systems amplify unlawful or harmful material
- that the best interests of the child are a primary consideration in the design and operation of services likely to be accessed by children
- that service providers make available controls that give end-users autonomy to support safe online interactions, and
- that service providers review and respond to reports and complaints within a reasonable period of time and provide feedback to users on the actions taken.

The Review could consider whether implementing a statutory duty of care model and an enforceable Safety by Design model is a sound approach to responding to these issues.

Question 5. Should the Act provide greater flexibility around industry codes, including who can draft codes and the harms that can be addressed? How can the code drafting process be improved?

The drafting process of the industry codes could be improved by including consultation with primarily affected communities (specifically regarding violent extremism and terrorism incidents).

Question 6. To what extent should online safety be managed through a service provider's terms of use?

Online service providers must play a significant role in managing harmful content, conduct and contact. Accountability should be enforced in the most efficient way possible, primarily via regulation.

Question 7. Should regulatory obligations depend on a service provider's risk or reach?

NSW suggests that obligations to ensure online safety should be similar across providers. Penalties for non-compliance should take into account the provider's risk, reach and turnover.

3.2 Protecting those who have experienced or encountered online harms

Question 8. Are the thresholds that are set for each complaints scheme appropriate?

Bystanders should also be able to lodge complaints (see Question 14). For the adult cyber-abuse scheme, the threshold could be reconsidered to include volumetric/pile-on attacks, which are particularly prevalent within online sexual abuse.

Question 9. Are the complaints schemes accessible, easy to understand and effective for complainants?

Consideration should be given to how the complaints scheme and professional development and training opportunities can be better advertised/promoted in an appropriate way to groups at greatest risk of victimisation including:

- children and young people
- people from culturally and linguistically diverse backgrounds
- people living with disability or medical conditions
- Aboriginal and Torres Strait Islander peoples
- people who identify as LGBTQIA+
- people with religious beliefs, and
- older Australians.

Campaigns promoting the schemes and informing people of their rights could be co-designed with targeted groups to ensure that messaging will have the greatest impact. Evaluation of impact on

online behaviours and awareness of rights/complaints systems may be useful to ensure effectiveness.

However, NSW notes that a framework that relies too heavily on individual reporting can unfairly shift the burden to users. This approach should be considered in tandem with other interventions.

Question 10. Does more need to be done to make sure vulnerable Australians at the highest risk of abuse have access to corrective action through the Act?

NSW suggests the Review consider provisions for blocking access to/closing sites which have repeated corrective action requests (e.g. 'revenge porn' sites), as well as accounts and 'communities' that are repeatedly directed to remove content. NSW would also support further education and awareness about the powers of the eSafety Commissioner through the complaints-based schemes (see question 16).

Children and young people with disabilities should be considered as a target group for improved accessibility to corrective action.

NSW suggests that the various services who routinely support vulnerable people should also have access to information, resources, and training on how to support their clients if they experience online abuse. Vulnerability can bring complexity to people's lives, and vulnerable people may not have the skills, resources, or capacity to action a complaint without additional support.

Cultural safety of the process, materials, supports is also an important consideration, as is ensuring that vulnerable people in regional and remote areas have equal access. Given digital connection issues for remote Aboriginal communities in particular, NSW recommends working with Aboriginal communities on solutions for support to ensure that their access to corrective action is prioritised. It is also noted the overrepresentation in the justice system of people with suspected or diagnosed disability — both for people offending and those who are also victims. Again, accessible communications about corrective action is important.

The Issues Paper states that 87 per cent of complaints via the Online Content Scheme related to child sexual exploitation and abuse, child abuse or paedophile activity. Further consideration could be given to appropriate resourcing to ensure victims of online abuse receive any required psycho-social support needed as a consequence of this, in alignment with objectives of *National Plan to End Violence against Women and Children 2022–32*.

Question 11. Does the Commissioner have the right powers to address access to violent pornography?

NSW suggests the Review reconsider the definition of 'violent'. Much mainstream pornography presents violence as a norm, and that absence of consent can be turned around through violent and coercive sexual actions. The Review may wish to consider whether this content, if permitted at all, should be classified as Class 1 material, including coercive, non-physically violent actions.

Current mechanisms under the OSA to regulate pornography are through developing industry codes and through complaints and removal under the Online Content Scheme. However as outlined in the Issues Paper, children are still at higher risk than previously of being exposed to pornography (including violent pornography). The most common source for exposure is through social media sites, more than pornographic sites. The Review could consider a duty of care model as a remedy to this, so that the responsibility for protecting consumers from the harms of their product is placed on the provider, rather than burden being placed solely on parents to monitor this.

Question 12. What role should the Act play in helping to restrict children's access to age inappropriate content (including through the application of age assurance)?

NSW welcomes the Australian Government's announced trial of age assurance technologies to help restrict children's access to age-inappropriate content.

The drivers for introducing age verification for online services are diverse:

- tackling children's access to violent online pornography and misogynistic content, which exacerbates violence against women and children (including child sexual abuse)
- the mental health harms, including increased depression, suicide and loneliness associated with increased in online activity, including social media
- safety threats posed by cyberbullying, image-based sexual abuse, and contact from unknown people
- exposure to inaccurate, polarising and extreme content that can lead to radicalisation and violence

Age verification assurance would likely need to consider access to pornography as well as social media, given the breadth of harms.

Restricting children's access to age-inappropriate content is an essential primary prevention strategy in the government response to childhood sexual abuse and problematic and harmful sexual behaviour (**PHSB**) displayed by children and young people.[2] [3]There is a significant cost[4] to government in responding to downstream issues related to children accessing pornography.

NSW has developed Children First 2022-2031 and the prevention action strategy Talking About It, which outlines our multi-agency approach to preventing and responding to PHSB. These plans identify that early access to pornography is a contributing factor to PHSB.

Increasing access to technology and the sexualised nature of online material including pornography and other sexually explicit materials contributes to sexual violence. It can also enable technology assisted PHSB to occur including compulsive use of pornography, creating, sending or distributing explicit images and viewing explicit materials online.

NSW suggests that the Review consider that the onus should be on the website/content provider to ensure that there are restrictions in place to limit access to Class 2 content, and penalties should be in place for failure (and repeat failure) to comply. This could include increased security on young people's accounts with restrictions on who can contact them. This could be a scaled approach whereby the platforms with largest reach and/or more harmful content would bear higher penalties.

Of course, there are positive benefits from young people engaging online — and any access restrictions need to weigh up privacy considerations from verifying age, access to information and

---

[2] Problematic and harmful sexual behaviours (PHSB) are defined as: sexual behaviours by children and young people under the age of 18 years that fall outside the range of expected activity for a child's age and stage of development may be developmentally inappropriate, harmful towards self or others, or abusive towards another child, young person or adult.

[3] PHSB is a significant and increasing issue in contemporary Australian society. Recent findings from the landmark Australian Childhood Maltreatment Study indicate that more young people aged 16-24 have experienced childhood sexual abuse from another adolescent than an adult (18.2% vs 11.7% respectively; Mathews et. al 2024).

[4] In 2020, the Productivity Commission estimated the annual cost of mental health disorders and suicide as $200-220 billion. Child abuse contributes substantially to this 'crippling national burden' (Australian Childhood Maltreatment Study 2023).

connectedness to communities, and young people's civil liberties. Moreover, the trend is for users to live more of their lives in a social media space — for shopping, direct communication, education and fin-tech services.

Table 1: Wellbeing domains (internal NSW Government analysis)

Wellbeing Domains	Physical Health	Mental Health	Relationships & Community	Worldviews & Attitudes	Learning & Growth	Safety & Security
Examples of positive impact from social media	Increased participation through gamification, exposure to content on healthy habits and diet	Access to expanded support networks and information, comfort to share, promotion of help-seeking behaviour	Access to expanded support networks, access to identity-affirming content, access to diverse peer groups not available offline	Access to learning content, diverse perspectives and experiences	New opportunities, individualised learning content, idea exchange and discussion	Access to resources, opportunities for skills development, location sharing features on social media can be a safety tool
Examples of resulting benefits	• Motivation • Routine • Increased activity	• Improved mental health • Social connections	• Sense of belonging • Increased confidence • Positive relationships	• Inclusive of others • Diversity of thought	• Engaged at school • Ambitious	• Improved knowledge of safety and security
Examples of negative impact from social media	Excessive screen time, sedentary behaviour, disrupted healthy behaviours	Exposure to harmful content, negative comparisons to others, reinforcement of negative thinking, harassment, fear of missing out	Taking time away from relationships, adding stress and conflict	Exposure to harmful content, exposure to siloed, inaccurate or polarising content	Distraction, disruption to learning of self and others, aggressive behaviours	Exposure to unsafe relationships, exposure to harmful content, scams, and threats of physical or sexual violence
Examples of resulting harms	• Not enough sleep • Low quality sleep • Inactivity and reduced participation in physical activity • Eye strain	• Anxiety and Depression • Suicide • Eating disorders • Aggression • Low self-worth • Addiction	• Disengagement • Isolation • Cyberbullying • Loneliness • Participation fatigue	• Cyberbullying • Radicalisation • Objectification of others • Criminal behaviour	• Disengagement • Lack of motivation, focus, direction • Poor educational outcomes	• Cyberbullying • Data/identity theft • Financial extortion • Crime • Grooming • Stalking • Exploitation/ abuse

Consideration of any age-based restrictions should:

- take into account the findings of the eSafety Commissioner's Roadmap for Age Verification on the immaturity and risks associated with current age assurance technologies (including the experience in UK in terms of learnings and challenges)
- meaningfully involve young people as stakeholders in their development, in line with the eSafety Commissioner's best practice approach to engaging young people as partners in research development and decision-making
- be trialled with strong consideration of safety and privacy of users, including the security of any data collected for this purpose
- be accompanied by resources for parents/carers to support children to build critical thinking skills and have regular open conversations about online content.

NSW would welcome engagement on the development of an age assurance trial. NSW has successfully piloted age verification for the purchase of online alcohol in partnership with Tipple and Mastercard in 2023. Key elements of the pilot's success were:

- providing customers with a choice of their preferred identity provider (not just one option)
- the identity service provider only sharing confirmation with the service provider (e.g. Tipple) that the customer is over/under the required age (in this case 18+ years old).

This approach meant that customers have control over what is shared with whom; they did not need to overshare personal information nor have that information unnecessarily retained by the service provider.

Question 13. Does the Commissioner have sufficient powers to address social media posts that boast about crimes or is something more needed?

As mentioned in the Issues Paper, while the OSA provides powers to remove Class 1 material, concerns remain about the impact of people sharing material of their criminal behaviour.

Section 154K of the *Crimes Act 1900* (NSW) was amended in March 2024 to incorporate a new offence for 'performance crime'. This offence applies in connection with motor theft, breaking & entering, and when the offender disseminates material to advertise the act of participation or involvement in the offence. Further, section 154K prohibits sending, supplying, exhibiting, transmitting, or communicating the material through social media and other electronic methods. The role of this amendment is to combat posting and boasting crimes in particular with young offenders.

A penalty of two years' imprisonment applies to people who commit motor vehicle theft or break and enter offences and share material to advertise their involvement in the criminal behaviour (including on social media).

With the amendment being so recent, it is difficult to determine whether the changes have had any impact or whether any further legislation change, including to the OSA, is required.

However, the Review may want to consider whether it could regulate content that does not meet the level of criminal behaviour, that nonetheless has negative impacts on young people and school communities such as student fight videos and vandalism.

Question 14. Should the Act empower 'bystanders', or members of the general public who may not be directly affected by illegal or seriously harmful material, to report this material to the Commissioner?

NSW suggests that the Review explore empowering any person to report offensive or harmful information online to the eSafety Commissioner and not restrict this to persons directly involved or impacted. This would also allow law enforcement agencies to report this type of content directly to the eSafety Commissioner.

Reporting on illegal or harmful material will assist in protecting vulnerable members of the community. The ability to report by the general public may also reduce the number of instances of child exploitation (an often under reported phenomena) and the impact of online bullying.

Allowing wider reporting would expand the OSA's protective function. Given the nature of online material, there may be harm caused to an adult or young person, even before the 'victim' even becomes aware of the content.

In particular, NSW supports consideration being given to introducing the possibility of bystanders to report intimate image-based abuse. This is important given that victims are often not aware that their images or digitally manipulated images of them are being circulated on sites or social media channels set up to encourage sexual violence/misogyny (e.g. 'revenge porn' sites, group sites set up to share violent photoshopped images of women). However, expansion to include bystander reporting should be led by survivor voices and account for privacy concerns.

Question 15. Does the Commissioner have sufficient powers to address harmful material that depicts abhorrent violent conduct? Other than blocking access, what measures could eSafety take to reduce access to this material?

NSW understands that the Online Content Scheme provides the eSafety Commissioner with a range of formal compliance powers for dealing with Class 1 material. NSW understands that the *Criminal*

Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 has not yet led to any criminal arrests or penalties (but reportedly has an effective deterrent effect).[5]

While the eSafety Commissioner is committed to a cooperative approach and graduated actions against online harm, there are times when enforcement action is warranted. Should the Review consider that these powers cannot be clearly or easily exercised, NSW suggests exploration of further powers.

The Review may wish to consider a NSW example as an approach for exercising powers. The declaration regime in Division 4, Part 2 of the *State Emergency and Rescue Management Act 1989* (NSW) provides that the Premier has the power to declare that a state of emergency exists in NSW where they are satisfied that an emergency constitutes a significant and widespread danger to life or property in the State. The declaration authorises the Minister for Emergency Services (and other persons) to exercise extraordinary powers to expedite response and recovery operations for the period while the declaration is in force.

The Review may also wish to consider whether the scope of Subdivision H of Division 474 of the Criminal Code (which provides that it is an offence for a social media service provider, designated internet service provider or hosting service provider to fail to expeditiously remove abhorrent violent material from their service) should be expanded, for example to cover online content published by anyone, rather than only the perpetrators of violence or their accomplices.

The Review could also consider using mechanisms for closing down/blocking access to repeat offender sites/accounts/communities which receive multiple complaints for removal of abhorrent violent conduct material. As previously mentioned, this could be reinforced by a duty of care/safety by design model whereby a positive obligation is imposed on platforms to moderate content such that the mere fact that it appears online attracts a penalty.

Measures to reduce access to material could also include:

- splash screens to delay and possibly turn around people accessing content
- an opportunity to not proceed
- contact details for services which can be accessed to provide support
- a note to user that history of their access to these sites and material is being held – by Internet Service Provider (**ISP**) and that cache be held by ISP for a significant period
- legislation permitting an owner of a device (such as a parent) to have access to history of access to material/sites (by a child).

Question 16. What more could be done to promote the safety of Australians online, including through research, educational resources and awareness raising?

In the awareness raising space, ongoing and more direct/explicit messaging is needed, particularly for young people and potentially less 'tech-savvy' parts of the community (including older parents/carers of children).

Promotion among young people needs to be direct and accessible, especially for young people who may not have a trusted and reliable adult support person or carer to help them develop protective

[5] Australian Government Parliamentary Joint Committee on Law Enforcement, Report into Review of Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019, December 2021, available at Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019 (aph.gov.au)

behaviours. Educational environments are one useful forum, however, not all young people are not meaningfully engaged with formal education. Considerations here need to also include digital literacy levels and information sharing in accessible formats.

The eSafety Commissioner's content removal schemes should be promoted much more highly to the general public so there is greater awareness. The Review may also consider exploring stronger links between awareness campaigns and sexual violence/sexual assault services, to enable these services to distribute information on the eSafety Commissioner's services.

Further research could be undertaken into the impacts of pornography and attitudes to women and use of sexual violence.

In NSW, the NSWPF and other NSW government agencies provide the following educational and training resources to promote online safety which could be considered in the national context:

- Youth Commands Youth Engagement Officers (**YEOs**) deliver an AFP resource called "Think U Know" which aims to engage parents, carers, educators, and police in raising awareness and preventing online child sexual exploitation. YEOs deliver the package in schools which focuses on the legal, social, and ethical considerations for technology use to prevent incidents. Schools request 'Think U Know' to be delivered by YEOs and in 2023, 432 presentations were delivered to 42,880 students.
- NSWPF provides educational resources and promotes awareness of online safety to the community across all Police Area Commands and Police Districts in NSW. Materials utilise various platforms targeting trending and emerging crime in collaboration with the ReportCyber Working Group, and the National Cybercrime Prevention Network. These resources are disseminated statewide to all relevant stakeholders. Additionally, they are accessible to the public via the NSWPF website and NSWPF social media pages (Facebook, Instagram, Snapchat, TikTok, X, WeChat and Weibo).
- Resources include but are not limited to a focus on online safety, fraud and online fraud, cybercrime, protecting children online, crypto safety, deepfake AI scams, rental scams, sextortion, employment scams, holiday scams, password security, technology facilitated abuse, arrest warrant scams etc.
- ID Support NSW in DCS provides the Be Prepared Guidelines with practical information for preparing and protecting against identity theft, including how to maintain online privacy, passwords and devices.

NSW's 'digital community resilience' initiatives take a whole-of-community approach that recognises that people can play different roles in creating safer and more inclusive online communities: beyond victims, perpetrators and bystanders to online hate, there are also roles played by convenors, moderators, creators, influencers, allies and a range of other active roles that people can play in online communities. The need for community education and awareness around online safety therefore includes, but extends beyond, reporting online harms to also include digital literacy and understanding of online platforms, promoting positive online behaviours, creating effective positive online campaigns, effective online community and network building etc.

MNSW had adopted this approach in implementing a suite of 'digital community resilience' initiatives, including a review and refresh of Remove Hate from the Debate initiative, which aims to inspire and empower young people with the tools and techniques to address online hate in a safe and effective way, and through a strategic partnership between the MNSW COMPACT Program and the Digital Industry Group that is working to build an online capability for the state-wide COMPACT Alliance to more effectively respond to online and offline threats to community harmony. The partnership includes the creation of a new COMPACT Digital Youth Alliance and a series of online

masterclasses for COMPACT partners and young people to build community capacity to combat online hate.

3.3 Penalties, investigation and information gathering powers

Question 17. Does the Act need stronger investigation, information gathering and enforcement powers?

Noting the difficulties identified in the Issues Paper around identifying perpetrators by the limited information available from online service providers, the Review may wish to consider extending powers to enhance this.

In terms of notices to end-users, NSW suggests consideration of how these may need to be different where the end-user is a child.

Question 18. Are Australia's penalties adequate and if not, what forms should they take?

NSW suggests the Review consider how penalties could be proportionately based on the size and reach of a service provider. As presented in the Issues Paper, similar penalties overseas and other Australian penalties for corporations are often tied to a percentage turnover. NSW would encourage strengthening accountability of service providers, such as through penalties tied to service provider turnover.

The Issues Paper suggests that penalties may fail to strike a proper balance between various offences, e.g. the maximum penalty for failing to take down illegal material such as child sexual exploitation material or pro-terror material is the same as for failure to take down harmful but not unlawful material (such as child cyberbullying or adult cyber-abuse material).

Question 19. What more could be done to enforce action against service providers who do not comply, especially those based overseas?

Noting ongoing legal proceedings on this issue, the Review may want to consider further changes to the OSA following decisions in these cases, including thresholds for classifying Class 1 material and Abhorrent Violent Material.

As previously indicated, NSW welcomes consideration of stronger and enforceable prevention mechanisms so that content is prevented from being promulgated online from the start, such as by requiring content recommender systems to identify and restrict access to harmful content.

For example, the United Kingdom is currently looking at ways to compel social media companies to improve their algorithms to focus on child safety. The use of artificial intelligence to detect suspicious conversations is critical to identify attempts to groom children and potential screen capture offences ('capping') where children are coerced into performing live-streamed sexual acts. Case studies, even de-identified, should form the basis of any considerations as to the effectiveness of legislation, practice or policy.

Moreover, further international cooperation, such as demonstrated through Christchurch CALL and INHOPE may be used to better achieve outcomes on a wider array of harmful content. An international approach is critical when online activity is effectively borderless.

Question 20. Should the Commissioner have powers to impose other enforcement actions such as business disruption sanctions?

NSW understands the eSafety Commissioner has significant powers to:

- apply for a Federal Court order for a social media service to stop providing that service in Australia
- issue an app removal notice to a provider of an app distribution service (e.g. Apple (IOS Apple Store) and Google (Google Play Store)), requiring them to cease enabling end-users in Australia to download an app that facilitates the posting of 'Class 1 material' on a social media service
- issue a removal notice to a designated internet service provider (i.e. Optus, iiNet), requiring the removal of Class 1 material from the internet service
- issue a direction for a social media service provider to comply with the Social Media Services Online Safety Code (Class 1A and Class 1B Material), such as the Code's requirements for Tier 1 social media services to remove 'Class 1A material' and 'Class 1B material' as soon as reasonably practicable
- issue a blocking notice to an internet service provider (i.e. Optus, iiNet), requiring the provider to block access to material that promotes, incites, instructs in or depicts 'abhorrent violent conduct'.

It is unclear how often these powers have been used and how effective they are likely to be in the context of a non-compliant service provider, particularly if based overseas. As mentioned previously, if these powers are not sufficient, other mechanisms should be explored.

3.4 International approaches to address online harms

Question 21. Should the Act incorporate any of the international approaches identified above? If so, what should this look like?

NSW suggests that best interests of the child principle and Safety by Design principle could be incorporated.

Question 22. Should Australia place additional statutory duties on online services to make online services safer and minimise online harms?

NSW welcomes the Review's consideration of additional statutory duties on online services to make online services safer and minimise online harms. A statutory duty of care model should incorporate the 'best interests of the child' principle set out in the United Nations Convention on the Rights of the Child and include penalties for non-compliance.

Question 24. Should there be a mechanism in place to provide researchers and eSafety with access to data? Are there other things they should be allowed access to?

To improve governments' ability to understand and respond appropriately to emerging risks to the community, mechanisms for greater transparency around safety controls should be considered as a part of the Review.

Platform development drastically outpaces processes for peer-reviewed research. This makes it extremely challenging to assess impacts of, for example, social media platforms on vulnerable groups such as harmful developmental impacts on young people.

There is also a significant information imbalance between platforms and the public, researchers and government. Whistleblowing events have shown that platforms are able to leverage enormous troves of data to understand risks to online safety. The Review could consider mechanisms for researcher access to service/platform data similar to those included in the EU's Digital Services Act, to facilitate public-interest research that improves understanding of systemic online risks to individual and community safety.

3.5 Regulating the online environment, technology and environmental changes

Question 27. Should the Commissioner have powers to act against content targeting groups as well as individuals? What type of content would be regulated and how would this interact with the adult cyber-abuse and cyberbullying schemes?

NSW agrees that this should be considered by the Commonwealth.

This is a particular issue for Aboriginal people, and young people from minority or socially disadvantaged backgrounds. In some cases, it would intersect with cyberbullying, but the ability to deal with targeting of groups could reduce harm for vulnerable people (such as refugee children, who are often concentrated in particular NSW schools).

For Aboriginal people, social media platforms have become unsafe places because of the extent and lack of consequence for racially motivated hate speech. Racism online causes significant harm to Aboriginal people.

While platforms have rules against use of hate speech on their platform their analytical tools or moderation struggle to effectively implement those rules, both in terms of identifying hate speech and/or to understanding regional and race-specific contexts of racist hate speech. This is particularly evident in hate speech against Aboriginal people. Perversely, these tools and moderations unfairly punish those who call out or respond to hate speech.

The current OSA is silent on hate speech or racist content, and in particular the eSafety Commissioner's powers and functions to investigate complaints about such content.

The Review could also consider that the eSafety Commissioner:

- publish guidelines on hate speech that recognise various forms of hate speech against particular groups or regional forms of racist speech
- receive and investigate complaints about online hate and to issue penalties against platforms who fail to take effective action against it, and
- provide education on harms of online hate and steps people can take to report it.

More broadly, NSW is concerned about the proliferation of online hate that seeks to incite fear and division among communities, often along racial or religious lines. Online hate in this context is often directed towards racial or religious groups rather than individuals as such. Online hate targeting groups threatens the online safety of all individuals who identify with that group and has the effect of generating fear and division within and between communities. NSW considers the eSafety

Commissioner has an important role in building community capacity and community resilience to support safer online communities.

Question 28. What considerations are important in balancing innovation, privacy, security, and safety?

Australia's online safety regulatory framework should consider core cyber security principles such as secure-by-design. The Australian Government has an important role to play in creating an environment where cyber security protections are embedded into online systems and apps to keep information, services, and systems safe.

Secure-by-design principles can ensure cyber threats are considered and data security safeguards are built into systems early. This enables the protection of consumer data and privacy by embedding shields such as access controls.

User-centred design frameworks can enable the development of solutions that prioritise user privacy and security by involving users in the design process to understand their needs and concerns.

- having strong cybersecurity infrastructure is important in protecting consumer information and privacy. For example, implementing multi-factor authentication, encryption and back-ups improves security and reduces the risk of a data breach
- effective cyber security governance structures provide a strategic view of organisational cyber security and helps maintain accountability
- establishing an expert advisory group to assist in providing advice and recommendations on effective ways to manage privacy and cyber security
- leveraging existing bodies and collaborating across Australian jurisdictions to share information, guidance and best practices
- improving education and cyber security awareness training to enhance online safety.
- compliance to various policies and legislation to ensure that citizen data is safe and appropriately managed. For example, the NSW Cyber Security Policy, the ACSC Essential Eight, *State Records Act 1998*, *Privacy and Personal Information Protection Act 1998*, *Health Records and Information Privacy Act 2002*, and other frameworks.

This work should be promoted as part of the implementation of the 2023-2030 Australian Cyber Security Strategy, in particular Shield 1 – Strong businesses and citizens and Shield 2 – Safe Technology.

Other considerations include ethical data collection, transparency, and consent, underpinned by privacy-by-design principles.

Question 29. Should the Act address risks raised by specific technologies or remain technology neutral? How would the introduction of a statutory duty of care or Safety by Design obligations change your response?

The way Australians experience the online world will continue to evolve, redefining our understanding of risks and harms and uncovering fresh challenges, benefits and threats to individual wellbeing and broader social cohesion. Cyber-security risks, rapid developments in new and emerging technologies, geopolitical competition, and social cohesion and trust are all factors that will continue to influence and amplify trends in online technology and activity:

- **Platform technology hardware and software is rapidly evolving**: Spatial computing, extended reality (**XR**) and metaverse technologies are leading to increasingly immersive online experiences and further blurring the boundaries between virtual and physical spaces. Greater use of AI and user data libraries will enable greater personalisation and potentially more powerful filter bubbles on online platforms.
- **Increasing human and synthetic (bots) personalities on platforms**: Broader and deeper access to platforms by humans and escalating creation of fake accounts and bots are amplifying potential risks and harms. Bots have potential to cause the same harm of humans albeit on a grander scale and with heightened intensity, with current concerns focused on issues like harassment, disinformation and foreign interference.
- **Further blurring of social media and other online service information**: GenAI and developments in Large Language Models (**LLMs**) may blur the provenance of information. This may leave users potentially unable to distinguish authoritative and trustworthy government information and information coming from other platform sources as they use online tools.
- **Decentralisation and diversification of social media platforms**: A shift towards decentralised social media platforms (**DAOs**) could lower platform accountability and make content moderation more challenging. Decentralisation may also allow for people to create their own self-managed social media platforms, which is becoming more feasible due to the accessibility of no-code and low-code platforms.
- **Rapid advancements in AI and GenAI technologies**: Rapid advancements in GenAI and AI technologies has introduced new risks as well as amplifying existing ones. For example, amplifying risks associated with creation of harmful or illegal content at scale (e.g. deepfakes, misinformation, content that infringes copyright, etc).

In light of a rapidly evolving online environment, it is critical that Australia's online safety regulatory framework is not linked to specific technologies or solutions, however it must cover new emergent risks created by technologies (e.g., deepfakes). Further, when the existing scale and reach of large online services/platforms is considered alongside the emerging risk of proliferation and amplification of synthetically produced content, there is a clear case for more economically efficient regulation of online harm that focuses upstream, on the design of systems and processes that make up online services and platforms, rather than solely downstream on the content posted by users.

As part of the Review, serious consideration should be given to a general duty of care requirement towards users of online services and platforms, that creates an obligation to:

- exercise care in relation to user harm
- conduct regular risk assessment to a defined standard
- establish mitigating measures, and
- evaluate effectiveness of measures on an ongoing basis.

This approach would complement existing content-based rules, while recognising that software and business model decisions made by platforms and services materially affect the content seen and produced online.

The Cabinet Office

GPO Box 5341
Sydney NSW 2001
www.nsw.gov.au/the-cabinet-office

NSW GOVERNMENT