EXPLANATORY STATEMENT

Online Safety Act 2021

Online Safety (Basic Online Safety Expectations) Determination 2022

Issued by the Authority of the Minister for Communications, Urban Infrastructure, Cities and the Arts

Purpose

The purpose of the Online Safety (Basic Online Safety Expectations) Determination 2022 (the Determination) is to set out basic online safety expectations for social media services, relevant electronic services and designated internet services. Providers of these services are expected to take steps to meet the Expectations included in the Determination and protect Australians from unlawful and harmful material and activity that falls within the remit of the enabling legislation the *Online Safety Act 2021* (the Act), or impedes the online safety of Australians.

Section 45 of the Act provides that the Minister may, by legislative instrument, make a determination that sets out basic online safety expectations for social media services, relevant electronic services and designated internet services. Sections 49, 52, 56 and 59 of the Act provide the eSafety Commissioner (the Commissioner) with power to require that a class, or provider, of a social media service, relevant electronic service or designated internet service, prepare and provide reports to the Commissioner on the extent to which the provider complied with provisions in the Determination over a period of not less than 6 months. Reports will allow the Commissioner to hold service providers to account for the steps they are taking to keep Australians safe online.

It is not intended that the Commissioner prescribe specific steps for service providers to take to meet the expectations. The Determination itself also does not prescribe how expectations will be met. This is intended to provide the highest degree of flexibility for service providers to determine the most appropriate method of achieving the expectations.

Notwithstanding that the Determination provides flexibility for service providers, it does outline a number of examples of reasonable steps that could be taken within the sections of the Determination. Not all reasonable steps have to be taken by all service providers. Rather, they are intended to provide guidance to service providers.

In determining how to comply with the Determination, service providers will also have regard for written guidance on the Commissioner's website and will undertake efforts to consult with the Commissioner on appropriate steps.

The notes on the provisions of the Determination are set out at Attachment A.

The Determination is a legislative instrument for the purposes of the *Legislation Act 2003*.

Legislative authority

The Determination was issued by the Minister under Section 45 of the Act.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Section 45 of the Act provides that the Minister may, by legislative instrument, make a determination that sets out basic online safety expectations for social media services, relevant electronic services and designated internet services. In particular:

- subsection 45(1) of the Act provides that the Minister may determine the basic online safety expectations for a social media service
- subsection 45(2) of the Act states that the Minister may determine the basic online safety expectations for each relevant electronic service included in a class of relevant electronic services specified in the determination and
- Subsection 45(3) of the Act states that the Minister may determine the basic online safety expectations for each designated internet service included in a class of designated internet services specified in the determination.

Subsection 45(4) of the Act provides a determination under this section does not impose a duty that is enforceable by proceedings in a court.

Section 46 of the Act provides the core expectations that must be specified in a determination of basic online safety expectations.

The Act provides that the Commissioner may require services to provide reports on their compliance with the Determination. While the Act does not include penalties for not meeting the Expectations in the Determination, non-compliance with a reporting requirement may lead to civil penalties or other actions by the Commissioner. For example, the Commissioner may decide to publish statements about the extent to which services are meeting the expectations.

Scope of the Determination

The Determination is made under the Act and therefore is focused on the online safety of Australians. This means, for example, that the Commissioner may not use the Act's reporting requirements to request information about defamation proceedings, adherence to privacy legislation or copyright infringements which do not relate to online safety.

End-user privacy

The Determination is focused on the systems, policies and processes that service providers may employ to prevent harm and respond to harm when it occurs. It is not expected that a service divulge personal information about a particular end-user to demonstrate its adherence to the Determination. However, a service would report on ways that data, behavioural signals and trends are being captured, to prevent and address unlawful and harmful material or activity.

Reporting of in-confidence information

Where a particular service shares commercial-in-confidence features or information with the Commissioner for the purposes of demonstrating compliance with the Determination, this information would not normally be made public. However, the Basic Online Safety Expectations are intended to enhance transparency and accountability of service providers. Therefore, service providers are encouraged to make reports publicly available, or agree that the Commissioner may do so. In addition, if the Commissioner forms a view that a service provider is not complying with one or more of the expectations, the Commissioner may

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

prepare a statement to that effect. If there is a statement, the Commissioner will share this statement with the service provider, and may also publish it on their website, if it is considered appropriate.

Reporting on the Determination is not limited to new safety features. In the event that a service provider is required to report on actions undertaken to meet the Expectations, service providers are expected to report on existing measures and the effectiveness and impact of ongoing measures.

If a particular service provider or class of services has been issued a periodic or non-periodic reporting notice or determination under the Act, that service or class of services is required to report on actions taken to meet one or more specified applicable basic online safety expectations in the manner and form specified by the Commissioner. When a particular provision of this Determination does not apply to a service, that particular service provider should explain in its report why the provision does not apply.

Appeal

There is no penalty for not complying with the Expectations described in the Determination, however, service providers can be required to report to the Commissioner on compliance with the Determination and failure to report is subject to civil penalties of up to 500 penalty units. A service provider can apply to have a decision by the Commissioner to issue a periodic reporting notice (Section 49 of the Act) or a non-periodic reporting notice (Section 56 of the Act) internally reviewed and/or reviewed by the Administrative Appeals Tribunal (AAT).

Commissioner's role

The Commissioner will have regard for when a service provider demonstrates an effort and commitment to improving online safety for its users when determining if that particular service is compliant with the Determination. Service providers are expected to discuss if they have limits to their ability to meet different provisions of the Determination with the Commissioner.

The Commissioner will take a risk-based approach towards harmful and unlawful material in assessing whether service providers are taking reasonable steps to minimise the provision of certain material on their services.

The effect of paragraphs 46(1)(g),46(1)(h), 49(3)(a), 52(3)(a), 56(3)(a) and 59(3)(a) of the Act is that service providers will not be required to report on actions undertaken to comply with the Determination until six months after it commences.

Relationship to other legislation

In meeting the provisions of the Determination it is not expected that service providers disclose information or undertake activities that are prohibited by Part 13 or Part 15 of the *Telecommunications Act 1997* (as is provided for in Section 218 of the Act).

In meeting the provisions of the Determination, it is not expected that service providers disclose information or undertake activities that would make the service non-compliant with obligations under the *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018*.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

In meeting the provisions of the Determination it is not expected that service providers disclose information or undertake activities that would make the service non-compliant with its obligations under the *Privacy Act 1988* (Privacy Act). Additionally, the Commissioner is an 'agency' for the purposes of the Privacy Act and hence therefore is bound by the Privacy Act. Any personal information collected by the Commissioner would be appropriately protected and dealt with in accordance with the Privacy Act.

Consultation

Section 47 of the Act requires the Minister to undertake specified consultation prior to making a determination under Section 45 of the Act. The Minister fulfilled this by undertaking a period of public consultation on the Determination from 8 August 2021 to 12 November 2021.

The Minister conducted public consultation by inviting submissions on the proposed Determination through the release of a draft Determination and a consultation paper on the website of the Department of Infrastructure, Transport, Regional Development and Communications (the Department).

To support consultation with industry and civil society, the Department conducted two roundtable meetings – the first with representatives of 7 family and community groups and the second with representatives of 15 key industry groups. The Department also held 11 meetings with industry and civil society stakeholders who requested an opportunity to discuss their views on the Determination.

77 unique submissions were received from a range of private citizens, civil society groups, commercial organisations and industry bodies as well as 1072 emails expressing concern about the impacts of internet pornography.

The Minister has considered feedback received throughout the consultation period in finalising the Determination. Changes made to the Determination following consultation included:

- Clarifying that service providers are not expected to build systemic weakness or vulnerability into encrypted services while undertaking actions to comply with Section 8 of the Determination;
- A new note that provides the Commissioner may publish relevant guidance materials on the eSafety website following consultations that take place under Section 7 of the Determination;
- A new additional expectation that allows the Commissioner to ask service providers for reports about the performance of safety measures that have been announced publicly; and
- Clarifying that service providers are expected to communicate changes to policies and terms of use to users in plain and accessible language.

The Office of Best Practice Regulation has confirmed that the preparation of a Regulation Impact Statement is not necessary, as the regulatory burden associated with reporting on the Determination was considered in the Regulation Impact Statement for the Online Safety Act (RIS ID 25408).

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Commencement

The Determination commences on the day after it is registered on the Federal Register of Legislation.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Statement of Compatibility with Human Rights

Prepared in accordance with Part 3 of the Human Rights (Parliamentary Scrutiny) Act 2011.

Online Safety (Basic Online Safety Expectations) Determination 2022

The Online Safety (Basic Online Safety Expectations) Determination 2022 (the Determination) is compatible with the human rights and freedoms recognised or declared in the international instruments listed in Section 3 of the *Human Rights (Parliamentary Scrutiny) Act 2011*.

Overview of the Determination

The Determination was issued by the Australian Government under Section 45 of the *Online Safety Act 2021* (the Act).

The Determination sets out basic online safety expectations for social media services, relevant electronic services and designated internet services to take steps to protect the Australian community from unlawful and harmful activity and material online. When the Commissioner gives a notice requesting or requiring information, or makes a reporting determination, under Part 4 of the Act, services are expected to report against the provisions of the Determination as specified in the notice or reporting determination. Matters within scope of the Determination are consistent with the enabling legislation and the policy objective of keeping Australians safe online.

The Determination provides flexibility for service providers to uplift online safety practices in a way that works for them. A number of examples of reasonable steps that could be taken are included within the Determination to provide guidance to service providers about what actions could be taken that could lead to compliance with the provisions. These reasonable steps do not necessarily have to be taken in order for a service provider to comply with the Determination. When reporting on the provisions, service providers are expected to consult with the eSafety Commissioner (the Commissioner) and have regard to guidance that is issued.

The purpose of reporting against the expectations in the Determination is to boost the transparency of services and to provide the Commissioner with a tool to hold services to account for the steps they take to keep Australians safe online.

Human rights implications

The principal human rights that the Determination are also engaged by the Act. These are:

- The right to freedom of expression primarily contained in Article 19 of the *International Covenant on Civil and Political Rights* (the ICCPR), and also referred to in Articles 12 and 13 of the *Convention on the Rights of the Child* (the CROC) and Article 21 of the *Convention on the Rights of Persons with Disabilities* (the CRPD);
- The prohibition on interference with privacy and attacks on reputation primarily contained in Article 17 of the ICCPR, and also referred to in Article 16 of the CROC, and Article 22 of the CRPD;

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

- The right to protection from exploitation, violence and abuse primarily contained in Article 20(2) of the ICCPR, and also referred to in Article 19(1) of the CROC and Article 16(1) of the CRPD; and
- The best interests of the child, contained in Article 3(1) of the CROC.

These rights, and how they are engaged in the Determination, are discussed below.

Freedom of expression

Rights relating to freedom of expression are recognised and protected by Article 19 of the ICCPR and by Articles 12 and 13 of the CROC.

Paragraph 1 of Article 19 of the ICCPR recognises that everyone shall have the right to hold opinions without interference. Paragraph 2 states that everyone shall have the right to freedom of expression. Paragraph 3 recognises that the exercise of the rights provided for in Paragraph 2 may be subject to certain restrictions. Paragraph 3 of that article as well as Paragraph 2 of Article 13 of the of the CROC limits the types of restrictions that may be imposed. Restrictions are as provided for by law and are necessary either in respect of the rights or reputations of others or for the protection of national security, public order, health and morals.

The Determination engages the right to freedom of expression insofar as it includes provisions for the minimisation of, and reporting of complaints about, certain material. The Determination outlines expectations that service providers take reasonable steps to minimise:

- Cyber-bullying material targeted at an Australian child;
- Cyber-abuse material targeted at an Australian adult;
- Non-consensual sharing of intimate images of a person (image-based abuse);
- Class 1 material under the Online Content Scheme;
- Class 2 material under the Online Content Scheme (preventing access for children); and
- Material promoting, inciting, instructing in or depicting abhorrent violent conduct.

Cyber-bullying, cyber-abuse, image-based abuse and abhorrent violent conduct

The Australian public recognises the significant harms that can occur online and expects an appropriate regime to be enacted to prevent and minimise these harms. Often the primary goal of victims of cyber-bullying, cyber-abuse and image-based abuse is to have material removed as soon as possible. Provisions in the Determination that expect social media services, relevant electronic services and designated internet services to minimise this material allows that objective to be met. Similarly, the availability of material that promotes, incites, instructs in or depicts abhorrent violent conduct may cause significant harm to the Australian community. The Determination has provisions that expect service providers to minimise this material to meet the objective of limiting the exposure of Australians to this harmful material.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

The Determination potentially restricts the right to freedom of expression in relation to a person who provides cyber-bullying material, cyber-abuse material, non-consensually shares intimate images or provides material promoting, inciting, instructing in or depicting abhorrent violent conduct on a social media service, relevant electronic service or designated internet service.

It does this in a manner that allows service providers to determine how this material or activity will be dealt with in a way that is consistent with achieving the intended policy outcome of responding to the harmful or unlawful material or activity.

In any circumstances, the Determination does not prescribe the manner in which this harmful or unlawful material must be handled.

Class 1 and class 2 material

The Determination sets out that class 1 and class 2 material, as defined under the Online Content Scheme in the Act, is expected to be minimised on social media services, relevant electronic services and designated internet services. Class 1 material includes child sexual exploitation material and pro-terrorist content. Class 1 material, in many circumstances, would likely offend against the standards of morality, decency and propriety generally accepted by reasonable adults (i.e. the material is, or would likely be, refused classification under the National Classification Scheme). Class 2 material is material that is, or would likely be, classified as X18+ (or, in the case of publications, category 2 restricted) or R18+ (or, in the case of publications, category 1 restricted) under the National Classification Scheme. Class 2 material is considered inappropriate for general public access and/or for children and young people under 18 years old. The Determination has provisions that expect service providers to minimise this material and, for class 2 material, to prevent its access by children.

Provisions that deal with class 1 and class 2 material in the Determination potentially restrict the right to freedom of expression for persons that create or share material that would be subject to a removal notice or remedial notice by the Commissioner under the Online Content Scheme. Removal of this material is consistent with the reasonable and proportionate objective of encouraging service providers to prevent unrestricted access to material that would be harmful to Australians, including children.

Protections of freedom of expression extended to the Determination

Exemptions for materials, as outlined Section 86 and 104 of the Act apply to the Determination. These factors ensure that the minimising of material or activity and complaints mechanisms to deal with such material or activity are reasonable and proportionate.

Section 233 of the Act provides that the Act does not apply to the extent (if any) that it would impinge the constitutional doctrine of implied freedom of political communications. This protection extends to the Determination and ensures that the Determination is consistent with the rights of freedom of expression as it relates to political communication.

Prohibition on interference with privacy and attacks on reputation

Paragraph 1 of Article 16 of the CROC recognises, among other things, the right of a child not to be subjected to unlawful interference with privacy or unlawful attacks on their honour

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

and reputation. Paragraph 2 recognises that children have the right to the protection of the law against such interference or attacks. Article 17 of the ICCPR and Article 22 of the CRPD contain similar rights.

The articles do not set out the reasons for which the guarantees in it may be limited, however, limitations contained in other articles, for example, those that are necessary in a democratic society in the interests of national security, public order, the protection of the rights or freedoms of others, might be legitimate objectives in appropriate circumstances. In any event, limitations on privacy must be authorised by law and must not be arbitrary.

The provisions of the Determination are directed towards protecting the preservation of privacy and reputation of vulnerable people. For example, the provisions at Paragraph 6(3)(b) provides that the most restrictive default privacy and safety settings be provided on a service or component of a service that is targeted at, or being used by, children.

To the extent that the Determination interferes with a person's privacy and reputation by outlining expectations around unlawful and harmful activity or material provided for by anonymous accounts and encrypted messaging services, these interferences are reasonable and proportionate. In relation to anonymous (and pseudonymous) accounts used by persons, the Determination expects that service providers take reasonable steps to prevent those accounts being used to deal with material, or for activity, that is unlawful or harmful. The Determination does not require identity verification processes, and does not prescribe actions to address this material or activity that would interfere with a person's privacy and correspondence. In relation to encrypted services, the Determination seeks that service providers detect and address material or activity on the service that is unlawful or harmful. Protections are included in the Determination to ensure that in meeting this expectation, service providers do not weaken encrypted messaging services.

The Determination does not require or expect service providers to undertake actions inconsistent with obligations under the *Privacy Act 1988*, the *Telecommunications Act 1997* or *Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018*. Any adherence to expectations around anonymous (or pseudonymous) accounts and encrypted services are not to conflict with obligations under a Commonwealth Act.

Additional protections of user privacy and reputation are provided in the circumstances where a social media service, relevant electronic service or designated internet service is not able to comply with provisions in the Determination. When a service provider cannot comply with an expectation, the service is expected to outline, in its reporting to the Commissioner, a clear explanation for why the expectation does not apply to them and why it is not addressed. It is ultimately a decision for the Commissioner as to whether an expectation applies to a service, and whether a provider is compliant.

The right to protection from exploitation, violence and abuse

The right to protection from exploitation, violence and abuse is primarily contained in Article 20(2) of the ICCPR and other related conventions. The ICCPR and related conventions requires Australia to take measures to protect persons from exploitation, violence and abuse.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

The Determination supports this right insofar as it sets out expectations that social media services, relevant electronic services and designated internet services protect people from the non-consensual sharing of intimate images, or abusive or bullying behaviour by minimising this type of material. Victims of high volume, cross-platform attacks (also known as volumetric or 'pile-on' attacks) are protected by this Determination to the extent that providers consult and cooperate with each other to detect these events and share information to help services prevent and deal with such material or activity.

The purpose of the Determination is to encourage service providers to take efforts to prevent these types of behaviour. To that extent, the Determination promotes the right to protection from exploitation, violence and abuse because services are expected to minimise abusive material or for persons and provide a way for persons to complain about that material.

Adherence to these provisions would result in this material and activity being prevented and addressed under a service provider's terms of use.

The best interests of the child

Article 3(1) of the CROC provides that in all actions concerning children, the best interests of the child shall be the primary consideration. The principle requires legislative, administrative and judicial bodies to take active measures to protect children's rights, promote their wellbeing and consider how children's rights and interests are or will be affected by their decisions and actions.

The Determination supports the best interests of the child by including provisions that provide guidance to social media services, relevant electronic services and designated internet services to ensure default privacy and safety settings on children's services. Provisions in the Determination expect service providers to ensure that the default privacy and safety settings of children's services (a service or a component of a service that is targeted at, or being used by, children) are set at the most restrictive level. Additionally, reasonable steps that offer guidance for service providers suggest that services could ensure that assessments of safety risks and impacts are undertaken provide additional protections for children online. This reasonable step would support the best interests of the child because it could result in services identifying and addressing gaps where children are exposed to harm online.

The best interests of the child are also supported by provisions that expect service providers to take reasonable steps to prevent access by children to class 2 material. This type of material, such as material that would be X 18+ content, may be harmful to children.

Conclusion

The Determination is compatible with the human rights and freedoms recognised or declared in the international instruments listed in Section 3 of the *Human Rights (Parliamentary Scrutiny) Act 2011*. The measures in the Determination promote the right to protection from exploitation, violence and abuse and the best interests of the child.

To the extent to which the measures in the Determination may engage with the right to freedom of expression and the prohibition on interference with privacy and attacks on reputation, any limitation is reasonable, necessary and proportionate to the goal of promoting and improving transparency and accountability of online services and improving online safety for Australians.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Attachment A

Notes on the Online Safety (Basic Online Safety Expectations) Determination 2022

Part 1 – Preliminary

Section 1 – Name

This section provides that the name of the Determination is the *Online Safety (Basic Online Safety Expectations) Determination 2022*.

Section 2 – Commencement

This section provides that each provision of the Determination is to commence on the day after the Determination is registered on the Federal Register of Legislation.

Section 3 – Authority

This section provides that the Determination is made under the *Online Safety Act 2021*.

Section 4 – Definitions

This section provides that in this instrument Act means the *Online Safety Act 2021*.

Part 2 – Basic online safety expectations

Division 1 – Purpose of this Part

Section 5 – Purpose of this Part

This section provides that the Determination establishes basic online safety expectations for social media services, any kind of relevant electronic service and any kind of designated internet service. The definition of a social media service is at Section 13 of the Act. The definition of a relevant electronic service is at Section 13A of the Act. The definition of a designated internet service is at Section 14 of the Act.

The note in this section provides that the core expectations in Part 4 of the Act are outlined in this Determination.

The Determination applies broadly to social media services, relevant electronic services and designated internet services. This is intended to reduce complexity for Australian end-users who may be seeking information about what service providers are expected to do to keep Australians safe online. Not all expectations in the Determination apply to every service provider. For example:

- A service provider might not offer a particular service (such as an encrypted messaging service or anonymous accounts).
- In undertaking actions to comply with a particular expectation, a service provider is not expected to breach obligations under other Commonwealth, state or territory legislation such as Parts 13 or 15 of the *Telecommunications Act 1997*.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

When a particular expectation does not apply to a service, the service provider should outline, in its report, a clear explanation for why that expectation does not apply to the services they offer.

Division 2 – Expectations regarding safe use

Section 6 – Expectations – Provider will take reasonable steps to ensure safe use

Subsection 6(1) is a core expectation under Paragraph 46(1)(a) of the Act. It provides that services subject to the Determination will take reasonable steps to ensure that end-users are able to use the service in a safe manner. The intention of this subsection is to uplift how services develop and implement products, policies and terms in a way that has regard for the safety of Australian end-users.

Subsection 6(2) is an additional expectation. It provides that services subject to the Determination will take reasonable steps to proactively minimise the extent to which material or activity on the service is unlawful or harmful.

'Unlawful' is used to refer to material or activity that is not permitted under a Commonwealth Act. For the purposes of this Determination, the term 'unlawful' refers to illegal material or activity dealt with under the Act and other unlawful material or activities that may have a negative impact on the online safety of Australians.

Harmful is used to mean material or activity that is not unlawful, but is covered within the scope of the Act (for instance, cyber-bullying material and class 2 material). It is also used to mean material or activity that should fall under a service provider's terms of use, policies and procedures and standards of conduct for end-users (as outlined in Section 14 of this Determination).

Services may meet this expectation by having effective systems and processes in place to prevent, detect and take action to address harmful or unlawful material or activity when it surfaces.

Subsection 6(2) does not expect service providers to anticipate all harms on its services before they occur. Rather, service providers should take proactive and reasonable steps to identify and address existing and emerging harms online.

Subsection 6(3) provides examples of reasonable steps that could be taken to guide service providers on what actions they could choose to undertake that would enable them to meet the expectations outlined in Subsection 6(1) and Subsection 6(2). The list under Subsection 6(3) is not exhaustive, and service providers may elect to take different steps to meet the expectations in a way that best suits their circumstances.

Paragraph 6(3)(a) suggests that service providers could meet the expectations in subsections 6(1) and 6(2) by developing and implementing processes to detect, moderate, report and remove (as applicable) material or activity on the service that is unlawful or harmful. In practice, this means that service providers could have moderation, reporting and removal systems to address unlawful or harmful material or conduct directed at their users.

Paragraph 6(3)(b) suggests that service providers could meet the expectations in subsections 6(1) and 6(2) by ensuring that default privacy and safety settings of a service that is targeted

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

at, or being used by, children are set at the most restrictive level. This example of a reasonable step that could be undertaken is purposefully flexible to allow the provider of a service to determine, in consultation with the Commissioner (under Section 7), what a 'most restrictive' level means for their service. The intent of this reasonable step is to protect children from harm.

Paragraph 6(3)(c) suggests that service providers could meet the expectations in subsections 6(1) and 6(2) by ensuring that their employees or contractors are trained in, and expected to implement and promote, online safety. For instance, a service provider could provide its staff with online or in-person internal or external online safety training and detail the outcomes of this training. The intent of this reasonable step is to equip employees of services to recognise and respond to online safety risks.

Paragraph 6(3)(d) suggests that service providers could meet the expectations in subsections 6(1) and 6(2) by continually improving technology and practices relating to the safety of end-users. This means that providers are routinely evaluating and enhancing the effectiveness of safety interventions, and, as new harms emerge online that were not anticipated in regulation, that service providers are able to take proactive steps to minimise end-users' exposure to the content and activity because their technologies and practices are up-to-date.

Paragraph 6(3)(e) suggests that service providers could meet the expectations in subsections 6(1) and 6(2) by ensuring that assessments of safety risks and impacts are undertaken, and safety review processes are implemented, throughout the design, development, deployment and post-deployment stages of the service. In practice, this means that services could take steps to identify how its systems, processes and policies are working to minimise unlawful and harmful material and activity. Assessments should also identify gaps in these systems and processes so that providers can take reasonable steps to address them.

Section 7 – Expectations – provider will consult with Commissioner and refer to Commissioner's guidance in determining reasonable steps to ensure safe use

Subsection 7(1) is a core expectation under Paragraph 46(1)(b) of the Act. It provides that in determining what reasonable steps are for the purposes of Subsection 6(1), the provider of the service will consult the Commissioner. Subsection 7(1) is intended to have the effect of establishing a dialogue between the Commissioner and service providers, and offers service providers the opportunity to outline the limitations to actions they can undertake to ensure safe use. For example, when reasonable steps outlined in subsection 6(3) are not appropriate for a service, that service provider may consult the Commissioner about alternative steps that could be taken to ensure safe use. It also establishes a means for information sharing so that the Commissioner and industry can share industry and social developments to improve online safety outcomes.

Subsection 7(2) is an additional expectation. It provides that, in determining what are reasonable steps for the purposes of complying with Subsection 6(1), a service provider will have regard to any relevant guidance material made available by the Commissioner. The Commissioner will provide guidance on how it will determine compliance with expectations on the eSafety website. Following registration of the Determination, the Commissioner will undertake consultation with service providers before publishing guidance. This will mean that guidance is based on evidence.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

The note to this subsection provides the Commissioner with discretion to publish specific guidance issued to all service providers, which may include information disclosed to it under subsection 7(2). However, the Commissioner will not normally include information that is commercial-in-confidence or to which the disclosing provider does not consent to being published. There may be circumstances where the Commissioner may be required to disclose information such as in response to requests made under the *Freedom of Information Act 1982*. The intent of providing the Commissioner with discretion to publish advice is to ensure transparency and consistency around the advice being offered to service providers.

Section 8 – Additional expectation – provider will take reasonable steps regarding encrypted services

Subsection 8(1) is an additional expectation. It provides that if a 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 unlawful or harmful.

'Encrypted services' is used to refer to services that translate data into another form, or code, so that only people or entities with access to a secret key or password can read it. One type of encrypted service is end-to-end encrypted messaging services that allows only users who are communicating with one another to read messages.

In meeting this additional expectation, service providers are not expected to monitor users' private communications. Rather, service providers should have systems, policies and processes in place to prevent harm and respond to harm when it occurs on encrypted services. For example, a service provider may respond to unlawful material on encrypted messaging services by using behavioural signals to identify accounts sharing unlawful materials. A service provider may also have processes in place to prevent banned users from establishing a new account.

Subsection 8(2) is an additional expectation. It limits the types of reasonable steps that service providers would undertake in complying with Subsection 8(1) to be steps that do not:

- Implement or build a systematic weakness or systematic vulnerability into a form of encrypted service (Paragraph 8(2)(a));
- Build a new decryption capability in relation to encrypted services (Paragraph 8(2)(b)); or
- Render methods of encryption less effective (Paragraph 8(2)(c)).

The intent of subsection 8(2) is to provide additional assurances to service providers that it is not expected that services access encrypted messages when taking steps to adhere to subsection 8(1).

Section 9 – Additional expectation – provider will take reasonable steps regarding anonymous accounts

Subsection 9(1) is an additional expectation. It provides that service providers allowing anonymous (or pseudonymous) accounts are expected to take reasonable steps to prevent those accounts being used to deal with material, or for activity, that is unlawful or harmful.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

‘Anonymous accounts’ and ‘pseudonymous accounts’ mean accounts that hide or disguise the identity of an end-user. Examples include using a false name (or a pseudonym), or a fake profile.

This means that service providers are expected to have measures in place to prevent their users from evading enforcement action by registering or using a new account under a different name to continue to cause harm. In meeting this expectation, service providers might also have processes in place to prevent, identify and address situations when an end-user is using multiple anonymous or pseudonymous accounts to post unlawful or harmful content or conduct unlawful or harmful activity, for example as part of an effort targeted at harming or overwhelming another end-user.

Subsection 9(2) outlines examples of reasonable steps that could be taken to guide service providers on what actions they could choose to undertake to meet the expectations outlined in Subsection 9(1). The list under Subsection 9(2) is not exhaustive, and service providers may elect to take different steps to meet the expectations in a way that best suits their needs.

Paragraph 9(2)(a) outlines reasonable steps that could be taken to ensure that service providers have processes that can prevent the same person from repeatedly using anonymous (or pseudonymous) accounts to post material. For example, providers could have processes that utilise web identifiers (cookies, IP addresses, browser fingerprinting), device or hardware identifiers, and other identifiers (such as account/behavioural analysis, metadata and traffic signals) to identify and stop re-registrations or alternate accounts in appropriate circumstances.

Paragraph 9(2)(b) outlines a reasonable step could be to have processes that require verification of identity or ownership of accounts. This step does not require service providers to employ a ‘real-name’ policy or otherwise ‘unmask’ users’ identity (i.e. service providers can preserve the confidentiality and privacy of users’ identifying information).

Section 10 – Additional expectation – provider will consult and cooperate with other service providers to promote safe use

Subsection 10(1) is an additional expectation. It provides that a service provider will take reasonable steps to consult and cooperate with other service providers to promote the ability of end-users to use all of those services in a safe manner. This means that service providers may be asked how they cooperate with other members of industry to identify new harms, trends and issues that may impact the safety of end users. The intent of this information sharing is to then allow service providers to prevent and deal with harmful and unlawful activity in a manner that suits their circumstances.

It is not expected that service providers would consult and cooperate in a way that diminishes a service’s intellectual property or expects sharing of commercial-in-confidence information. Rather, the intent is that service providers consult and cooperate in relation to preventing and dealing with unlawful and harmful material or activity that adversely impacts online safety for Australians.

Subsection 10(2) provides examples of reasonable steps that could be taken to guide service providers on what actions they could choose to undertake that would enable them to meet the expectations outlined in Subsection 10(1). The list under Subsection 10(2) is not exhaustive,

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

and service providers may elect to take different steps to meet the expectations in a way that best suits their circumstances.

Paragraph 10(2)(a) suggests that a reasonable step that a service could take to cooperate with other services is to detect high volume, cross-platform attacks (also known as volumetric or 'pile-on' attacks).

High volume, cross-platform attacks occur when a person is named in, tagged, or linked to an abusive post, which others 'like', share, re-post with additional commentary, and/or link to via other services. The volume of material can proliferate rapidly across platforms.

Cooperating to promote safe use in this way could include making other services aware of a volumetric attack by sharing information like URLs, hashtags or account names. This information, in turn, could then assist a service to respond.

Paragraph 10(2)(b) suggests that a reasonable step could be to share information with other service providers about material or activity that is unlawful or harmful with a view to preventing and dealing with it. For example, services could share information about a subsection of the community that may be being targeted with abuse due to an identifying characteristic (like sexuality, ethnicity or disability). Services that receive this information could then take appropriate actions to prevent and deal with harmful material or activity targeted at that group.

Division 3 – Expectations regarding certain material and activity

Section 11 – Core expectation – provider will take reasonable steps to minimise provision of certain material.

This is a core expectation under Paragraph 46(1)(c) of the Act. It provides that service providers are expected to take reasonable steps to minimise the provision of certain material that falls within the Commissioner's remit under the Act.

Paragraph (a) provides that cyber-bullying material targeted at an Australian child, as defined in Section 6 of the Act, is subject to this expectation, and should be minimised on services.

Paragraph (b) provides that cyber-abuse material targeted at an Australian adult, as defined in Section 7 of the Act, is subject to this expectation, and should be minimised on services.

Paragraph (c) provides that non-consensually shared intimate images of a person, as defined in Section 15 and 16 of the Act, are subject to this expectation and should be minimised on services.

Paragraph (d) provides that class 1 material, as defined in Section 106 of the Act, is subject to this expectation and should be minimised on services. In determining what constitutes class 1 material, services may have regard for regulatory guidance published on the Commissioner's website. An example of class 1 material is livestreamed child sexual abuse material whether or not the material depicts abuse occurring in Australia.

Paragraphs (e), (f), (g) and (h) provide that material that promotes, incites, instructs in or depicts abhorrent violent conduct, is subject to this expectation and should be minimised on services. Abhorrent violent conduct has the same meaning as in subdivision H of Division 474 of the *Criminal Code Act 1995*. Service providers are not expected to minimise the

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

provision of material that is exempt material under Section 104 of the Act such as material produced for public interest journalism or advocacy for a change in legislation.

Section 12 – Core expectation – provider will take reasonable steps to prevent access by children to class 2 material

Subsection 12(1) is a core expectation under Paragraph 46(1)(d) of the Act. It provides that services will take reasonable steps to ensure that technological or other measures are in effect to prevent access by children to class 2 material provided on the services. The definition of class 2 material is provided in Section 107 of the Act.

Technological or other measures are the systems, policies and processes that service providers have in place, or may employ, to prevent harm and respond to harm when it occurs.

Subsection 12(2) provides examples of reasonable steps to guide service providers on what actions they could choose to undertake that would enable them to meet the expectations outlined in Subsection 12(1). The list under Subsection 12(2) is not exhaustive, and service providers may elect to take different steps to meet the expectations in a way that best suits their circumstances.

Paragraph (a) suggests that a service could meet the expectation under Subsection 12(1) by implementing age assurance mechanisms. An example of an age assurance mechanism is a Restricted Access System, which is outlined in Section 108 of the Act.

Paragraph (b) suggests that a service could meet the expectation under Subsection 12(1) by conducting child safety risk assessments. In practice, this would mean that services would outline what child safety risk assessments are undertaken by the service, what findings were made and what actions were taken. A child safety risk assessment may include determining if a child can access class 2 material on the service or whether a child using the service could be contacted by someone unknown to them.

Division 4 – Expectations regarding certain material and activity

Section 13 – Expectations – provider will ensure mechanisms to report and make complaints about certain material

Subsection 13(1) is a core expectation under Paragraph 46(1)(e) of the Act. It provides that services will ensure that they have clear and readily identifiable mechanisms that enable end-users to report and make complaints about any of the material in paragraphs (a) to (i). Paragraphs (a) to (i) cover material that is subject to actions by the Commissioner under the Act.

Subsection 13(2) is an additional expectation. It provides that services will ensure that they have clear and readily identifiable mechanisms that enable persons ordinarily resident in Australia to report and make complaints about any of the material in paragraphs (a) to (i). Paragraphs (a) to (i) cover material that is subject to actions by the Commissioner under the Act.

The intention of this section is to make sure that services have appropriate complaints processes for all Australians to use to report unlawful or harmful material regulated under the Act to a service without the requirement to have an account with a particular service.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Section 14 – Additional expectation – provider will ensure service has terms of use, certain policies etc

This is an additional expectation. It provides that the provider of a service will ensure that that service has terms of use and certain policies available to Australian end-users.

Paragraph (a) provides that the provider of a service has terms of use. Terms of use are a mechanism for service providers to outline what is and is not allowed on their service. Appropriate terms of use should address material that is subject to the Act and other online harms.

Paragraph (b) provides that the provider of a service has policies and procedures in relation to the safety of end-users. This paragraph is drafted broadly, but should include policies to address safety risks associated with unlawful or harmful activity or material.

Paragraph (c) provides that the provider of a service has policies and procedures for dealing with reports and complaints mentioned in Section 13 or 15 of the Determination.

Paragraph (d) provides that services have standards of conduct for end-users (including in relation to material that may be posted using the service by end-users, if applicable), and policies and procedures in relation to the moderation of conduct and enforcement of standards. This means, in practice, the service provider would have effective systems in place to enforce its own terms of use and policies.

Subsection 14(2) is an additional expectation. It provides that the provider of a service will take reasonable steps to ensure that penalties for breaches of its terms of use are enforced against all accounts held or created by an end-user who breaches the terms of use of that particular service. For instance, if an end-user was suspended from a service for breaching its terms of use, that service provider could be expected to have systems and processes in place to prevent the suspended end-user from creating an anonymous or pseudonymous account in order to post unlawful or harmful content, or conduct unlawful or harmful activity.

Services should use their terms, policies and procedures to address harmful material that is not necessarily unlawful or explicitly referenced in the Act, for example:

- Hate against a person or group of people on the basis of race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, gender identity, serious disease, disability, asylum seeker/refugee status, or age;
- Promotion of suicide and self-harm content, such as pro-anorexia content, that does not meet the threshold of class 1 or class 2 material;
- High volume, cross-platform attacks that have a cumulative effect that is damaging but does not meet the threshold of adult cyber-abuse when reported as singular comments or posts; and
- Promotion of dangerous viral activities that have the potential to result in real injury or death.

Subject to consultation with the Commissioner, when a service has terms of use, policies and procedures and standards of conduct that deal with unlawful and harmful activities or materials, they may be asked to report on these matters.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

The first note under this section refers to Section 17 of the Determination, which provides more detail on the information that should be accessible to end-users, and therefore accounted for under a service provider's policies.

The second note states that for Paragraph (b), the policies and procedures might deal with the protection, use and selling of end-users' personal information.

Section 15 – Expectation – provider will ensure service has mechanisms to report and make complaints about breach of terms of use

Subsection 15(1) is a core expectation under Paragraph 46(1)(f) of the Act. It provides that a provider of a service will ensure that the service has clear and readily identifiable mechanisms that enable end-users to report, and make complaints about breaches of the service's terms of use, as they are outlined in Paragraph 14(a).

Subsection 15(2) is an additional expectation. It provides that services will ensure that there are clear and readily identifiable mechanisms that enable persons ordinarily resident in Australia to report, and make complaints about, breaches of the service's terms of use.

The purpose of this section is to provide an avenue for all Australians to have material or activity that breaches a service's terms of use removed or otherwise dealt with in an appropriate manner by the service without the requirement for a user to have an account with a particular service.

It is not expected that a service provider will remove all material that is reported to it. This section would not expect a service to respond to complaints that are vexatious and do not breach terms of use. Instead, a service should take action to respond to breaches of terms of use.

Section 16 – Provider will make accessible information on how to complain to Commissioner

This is an additional expectation. It provides that the provider of a service will ensure that information and guidance on how to make a complaint to the Commissioner, under the Act, about material mentioned in Section 13 of the Determination is provided on the service, and is readily accessible to end-users.

The purpose of this expectation is to make Australian end-users aware that they can make complaints to the Commissioner, in response to material mentioned in Section 13 of the Determination. It is at the discretion of service providers to decide how they provide this information. Service providers may choose to make this information accessible at all points of the end-user experience, at the point of account creation, first use, regular intervals or in a sequence appropriate for that services' complaints process.

Division 5 – Expectations regarding making certain information accessible

Section 17 – Additional expectation – provider will make information on terms of use, policies and complaints etc. accessible

This is an additional expectation. It provides for the release, accessibility, review and presentation of information regarding a service's terms of use and information provided for by Section 14 (Paragraph 17(2)(a)), and information regarding online safety and parental

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

control settings – including in relation to the availability of tools and resources published by the Commissioner (Paragraph 17(2)(b)). At a minimum, this information should be provided at initial 'sign up' of accounts.

Information provided for in Section 14 (Paragraphs 17(2)(a) and 17(2)(b)) is expected to be provided in such a way that is readily accessible to end-users (Paragraph 17(1)(a)) and written in plain language. The purpose of this provision is to ensure that the end-users of a service can readily understand the services' terms of use, policies and complaints mechanisms.

Information provided for in Paragraph 17(2)(b) is expected to be accessible at all points in the end-user experience, including, but not limited to, point of purchase, registration, account creation, first use and at regular intervals. The purpose of this provision is to make it as simple and easy as possible for users to locate and make use of tools and settings to make their (or their children's) user experience as safe and age-appropriate as possible. This is particularly important when a user is registering for a service or using the service for the first time, but it is also important that the information be easy to find throughout a user's experience of the service. For the purpose of Paragraph 17(1)(b), provision of this information at 'regular intervals' would be satisfied through adhering to the expectations set out in section 18.

Section 18 – Additional expectation – provider will provide updates about changes in policies, terms and conditions etc.

This is an additional expectation. It provides that service providers should ensure that their end-users receive updates in plain language regarding changes in the information specified in Paragraphs 17(2)(a) and 17(2)(b), including through targeted in-service communications. This information is to be presented in plain language.

The purpose of this section is to ensure that Australian end-users are made aware of changes to a service's terms of use, policies, procedures, standards of conduct, parental controls and online safety settings, etc. – if and when they occur.

Division 6 – Expectations regarding record keeping

Section 19 – Additional expectation – provider will keep records regarding certain matters

This is an additional expectation. It provides for the retention by service providers of records of reports and complaints about the material mentioned in section 13, for a period of five years after the making of the report of complaint to which the record relates.

The purpose of this expectation is to allow services to provide the Commissioner with information about complaints being made to, and how they are actioned by, service providers. Over time, this information will help the Commissioner assess the effectiveness of complaints and moderation practices over time, and point out areas where services are doing this well, as well as areas where improvements could be made.

This expectation is not retrospective. Service providers are only expected to develop measures and processes to retain records of reports and complaints made after the making of

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

the Determination. Service providers will therefore not be expected to have five years of records until at least five years following the making of the Determination.

Division 7 – Expectations regarding dealings with the Commissioner

Section 20 – Expectations – provider will provide requested information to the Commissioner

Section 20 sets out three core expectations (contained in Paragraphs 46(1)(g), 46(1)(h) and 46(1)(i) of the Act) regarding a service provider's provision of information to the Commissioner. It also outlines one additional expectation relating to the reporting of announced safety features.

Under Subsection 20(1), the Commissioner may, by written notice to a provider of a service, request a statement that sets out the number of complaints made during a specified period to the provider about breaches of the service's terms of use. This information can serve several uses for the Commissioner. For example, it can provide the Commissioner with an indication of how well a service's terms of use are being provided to users, in line with expectations at Sections 17 and 18.

It is at the discretion of the service to provide additional information regarding these complaints (e.g. how many were deemed vexatious, how they were resolved etc.).

Under Subsection 20(2), the Commissioner may similarly request a statement that, for each removal notice given to the provider during a specified period, sets out how long it took a provider to comply with the removal notice.

In each subsection above, the specified period can be no shorter than six months, and the provider is given 30 days after the request is given to comply. This information will help the Commissioner assess how rapidly service providers are complying with removal notices given under the Act's schemes.

Under Subsection 20(3), the Commissioner may, by written notice given to the provider of the service, request that the provider give to the Commissioner specified information relating to the measures taken by the provider to ensure that end-users are able to use the service in a safe manner. The provider will be expected to comply with the request within 30 days after the notice of the request is given. The purpose of this expectation is to enable the Commissioner to request specified information concerning online safety measures being taken by a service provider.

Subsection 20(4) is an additional expectation. It provides that a service provider will, within 30 days of receipt of a written notice by the Commissioner, report on the performance of safety measures that it has publicly announced or reported to the Commissioner. In practice, this means that when a service provider announces a safety feature, that particular service provider can reasonably be expected to be asked by the Commissioner to report on the impact of that safety feature on the experience of its end-users. The intention behind Subsection 20(4) is to reduce instances of service providers announcing safety features, but then failing to disclose whether those features have worked or not.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062

Section 21 – Additional expectations – provider will have designated contact point

This is an additional expectation. It provides for the creation by the service provider of a designated point of contact for the purposes of communication with the Commissioner regarding the enforcement of the Act. It is important that the Commissioner is able to make contact with a service provider in order to better ensure timely compliance with the provisions of the Act, such as when the Commissioner issues a removal notice to have a non-consensually shared intimate image taken down.

Service providers are expected to have an email address and voice contact designated as the service's point of contact for the purposes of the Act. The provider will ensure these contact details are notified to the Commissioner, and that any change to these details will be notified in writing to the Commissioner within 14 days of the change.

Authorised Version Explanatory Statement registered 23/01/2022 to F2022L00062