Draft Online Safety (Basic Online Safety Expectations) Determination 2021

Consultation Paper

July 2021

Purpose

On 23 June 2021, the Australian Parliament passed the *Online Safety Act 2021* (the Act). The Act will commence on 23 January 2022. A key principle underlying the Act is that the rules and protections we enjoy offline should also apply online. The Act creates a modern regulatory framework that builds on and strengthens existing arrangements and holds industry to greater account for the safety of their users.

Online safety is a shared responsibility and industry has a unique and important role to play in supporting safer online spaces. Under the Act, the Minister for Communications, Urban Infrastructure, Cities and the Arts has the power to determine, by legislative instrument, basic online safety expectations for certain online services (Part 4 of the Act). This legislative instrument provides a mechanism for the Government to articulate its expectations of online service providers, on behalf of the community, to improve protections for users.

The Minister is now consulting on a draft *Online Safety (Basic Online Safety Expectations) Determination 2021*. The draft Determination includes the Government's core expectations, which are specified in the Act, as well as a number of additional expectations. Service providers are expected to take reasonable steps to meet these expectations in order to uphold the safety of Australian end-users on their services. The eSafety Commissioner also has the power to require service providers to report on their compliance with these expectations.

The draft Determination is available on the Department of Infrastructure, Transport, Regional Development and Communications website. We are inviting service providers directly affected by the provisions in the draft Determination, and other interested stakeholders, to make submissions on the draft Determination.

Background

The Basic Online Safety Expectations (the Expectations) arose out of significant reform to Australia's online safety legislative framework. A 2018 review of Australia's online safety legislation (Briggs Review) recommended that a new Online Safety Act should be developed and that this Act should '*guarantee that the online industry goes beyond simple compliance with minimum safety standards and should establish a much higher new benchmark standard with which all industry must comply*.'

Until now, the Government's expectations of industry have been articulated by three means: the basic online safety requirements in the *Enhancing Online Safety Act 2015* (EOSA), the eSafety Commissioner's Safety-by-Design principles, and the Online Safety Charter. These did not have any legal or regulatory enforceability.

1. The EOSA included a limited set of 'basic online safety requirements' at section 21. These applied to social media services, and broadly required that the service's terms of use prohibited cyberbullying, the service had a complaints scheme where end-users could request the removal of cyberbullying material, and the service had a designated point of contact for the eSafety Commissioner.
2. Safety-by-Design is an internationally recognised initiative of the eSafety Commissioner that seeks to influence the way that technology is designed, developed and deployed, shifting the responsibility for safety of users back onto technology companies and service providers. The eSafety Commissioner worked with more than 60 organisations across industry, government and civil society to arrive at three overarching Safety by Design principles:
    - *2.1 Service provider responsibilities*
    - *2.2 User empowerment and autonomy*
    - *2.3 Transparency and accountability*
3. The Online Safety Charter was launched by the Minister on 11 December 2019. The Charter was the Government's articulation of community expectations of technology firms and digital platforms to protect citizens, especially children and vulnerable members of the community, from harmful online experiences.

While online service providers have instituted a number of reforms and initiatives in recent years to improve the online safety of their platforms, end-users are still regularly exposed to harmful online content and experiences. Currently, too much responsibility falls on the end-user to take care of their own safety online. Without a new set of legislated expectations around minimum standards for pre-emptive and preventative action from service providers, there is a risk that user safety measures will continue to be reactive, and the burden of safety will continue to fall disproportionately on the end-user.

Online Safety Act 2021

On 23 June 2021, Parliament passed the *Online Safety Act 2021* (the Act). Part 4 of the Act sets out the requirements for the determination of the basic online safety expectations (the Expectations) for social media services, relevant electronic services and designated internet services and the requirements for compliance reporting by industry.

The Expectations articulate the Government's minimum safety expectations of online service providers, establishing a benchmark for online service providers to take proactive steps to protect the community from abusive conduct and harmful content online. The Expectations do not prescribe how these expectations will be met. Indeed, they have been crafted in a way that allows flexibility in the method of achieving these expectations.

What are the core expectations?

Section 46 of the Act sets out core expectations which are intended to apply to all social media services and all members of the defined classes of designated internet services and relevant electronic services. A determination under section 45 must specify the core expectations at section 46. The core expectations are principles-based and intended to be read in their broadest sense.

Expectations regarding safe use

It is a core expectation that the provider of a service will take reasonable steps so that end-users are able to use the service in a safe manner (see paragraph 46(1)(a)). The service provider will be expected to consult with the eSafety Commissioner to determine what reasonable steps means for that provider (see paragraph 46(1)(b)). This could be by seeking the advice of the Commissioner or following guidance issued by the Commissioner. An example of how a service provider might do so is by updating the service's terms of use to include standards of end-user behaviour, or to prohibit cyber-bullying, cyber-abuse, the non-consensual sharing of intimate images, material that depicts, promotes, incites or instructs in abhorrent violent conduct, class 1 material and class 2 material, and other harmful material from the service.

'Other harmful material' is intended to capture emerging forms of harmful material and behaviours that are not already specified in the Act or Expectations. This may include, for example, an end-user posting an image or video of a second person that uses artificial intelligence technology to transpose the second person's head on a computer-generated body (known as a 'deepfake'), for the purpose of misrepresenting the second person. Such an image may invite cyber-bullying, cyber-abuse, or a volumetric attack against the second person. Service providers are best placed to identify these emerging forms of harmful end-user conduct or material, and so the flexibility of this regime means that providers can choose the best way to address them on their service in the most responsive way. The instrument provides examples of reasonable steps that could be taken to ensure safe use.

Expectations regarding certain material and activity

It is a core expectation that the provider of a service take reasonable steps to minimise the extent to which the following materials are provided on their services (see paragraph 46(1)(c)):

- cyber-bullying material targeted at an Australian child;
- cyber-abuse material targeted at an Australian adult;
- non-consensual intimate images of a person;
- class 1 material;
- material that promotes abhorrent violent conduct;
- material that incites abhorrent violent conduct;
- material that instructs in abhorrent violent conduct; and
- material that depicts abhorrent violent conduct.

For example, setting clear minimum standards for online behaviour relating to the posting of such material and establishing clear protocols and consequences for service violations, including account suspension, access restrictions and blocking repeat offenders, would be considered a reasonable step in meeting this expectation.

It is a core expectation that the provider of a service will take reasonable steps to ensure that technological or other measures are in effect to prevent access by children to class 2 material (defined at section 107 of the Act) provided on the service (see paragraph 46(1)(d)).

Expectations regarding reports and complaints

It is a core expectation that the 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, certain material (see paragraph 46(1)(e)) as well as similar mechanisms to report breaches of the service's terms of use (see paragraph 46(1)(f)). To meet this core expectation, complaints and reporting systems should be accessible, fair, responsive and effective in dealing with reports of harmful content and conduct.

Expectations about making certain information accessible & dealing with the Commissioner

It is a core expectation that the provider will report to the Commissioner within 30 days of receiving a request by the Commissioner seeking the number of complaints made to the provider regarding breaches of the service's terms of use within a set period (see paragraph 46(1)(g)); how long the provider took to respond to any removal notices issued to it by the Commissioner under the Act during a set period (see paragraph 46(1)(h)); and what measures the provider is taking to provide for end-users to use their service safely (see paragraph 46(1)(i)).

How services can demonstrate they are meeting the Expectations

Generally, service providers that can demonstrate they have the necessary capabilities (both technology and human), skills, processes, systems and scalable solutions to proactively detect and respond to online harms occurring on their service would likely meet these expectations. For example, embedding user safety considerations, training and practices into the roles, functions and working practices of those employed by the provider and putting in place preventative measures to detect, surface, hash (where applicable), flag and remove illegal and harmful conduct and content would also be considered a reasonable step in meeting the core expectations.

What are the reporting requirements?

Under the Act, the eSafety Commissioner may require a provider or class of providers to report on their compliance with one or more basic online safety expectations. Subdivision 3A (periodic reporting about compliance with the basic online safety expectations) and subdivision 3B (non-periodic reporting about compliance with the basic online safety expectations) provide the Commissioner with the power to issue:

- a periodic report notice requiring an individual provider to report to the Commissioner on their compliance with the basic online safety expectations multiple times at regular intervals;
- a periodic report determination requiring each provider within a class of providers to report multiple times at regular intervals;
- a non-periodic report notice requiring an individual provider to prepare only a single report to be given to the Commissioner;
- a non-periodic report determination requiring each provider within a class of providers to each prepare a report to be given to the Commissioner.

The Commissioner may seek information about a provider's compliance with *all* or *one or more specified* basic online safety expectations. The provider must prepare the report in the manner and form specified in the Commissioner's notice, and give the report to the Commissioner either within the time period specified in the report, or such longer period as the Commissioner allows (but not less than 28 days).

Penalties may apply for those that fail to report

A provider who fails to comply with such a notice from the Commissioner may be subject to a civil penalty. In addition to a court ordered civil penalty, the Commissioner has access to other enforcement options, including formal warnings, infringement notices, enforceable undertakings and injunctions.

The Commissioner is able to publish statements about how providers are meeting the Expectations

The Act also provides for the Commissioner to publish statements about the performance of service providers in meeting the Government's expectations. It is intended that the Expectations, reporting to the Commissioner and public statements will provide much needed transparency about the level of harm occurring on services used by Australians and help to drive improvements in online safety practices by industry.

How do the Basic Online Safety Expectations fit with industry codes of practice?

The Act maintains the co-regulatory approach to online content regulation which has been the approach in Australia since the late 1990s. That is, legislation underpins the development of industry codes of practice by industry. These codes are registered and enforced by the eSafety Commissioner. When codes are non-existent or deficient, the eSafety Commissioner may create industry standards.

The purpose of the Expectations and industry codes of practice

The Expectations and the industry codes both seek to achieve the same outcome of keeping people safe online. While there is potential overlap between the codes and the Expectations, they serve different regulatory purposes.

The purpose of the Expectations is to place greater responsibility on service providers to ensure they provide safer services to Australian end-users. The Expectations provide flexibility for service providers to meet these Expectations. This approach recognises that traditional regulation may not suit the way content is created and delivered to users today. An example of this flexible approach is the expectation that service providers do more to assess and anticipate risks of harm facilitated by their services and take proactive and preventative action or 'reasonable steps' to mitigate those risks. Service providers are required to report on their compliance with the Expectations.

The purpose of industry codes and standards is to set out binding self-regulatory procedures directed at ensuring class 1 and class 2 material is limited on services accessible to Australian end-users. The requirements for industry to develop codes of practice and adhere to industry standards are not new. These arrangements have been brought over from Schedules 5 and 7 of the *Broadcasting Services Act 1992* and now fall under Part 9 of the Act (the online content scheme). The industry codes are likely to deal with more specific matters than the Expectations. Examples of matters that *may* be dealt with by industry codes and industry standards include procedures for responding to complaints and notices, the provision of certain information and advice by service providers and technological tools and solutions that must be offered by service providers to deal with class 1 and class 2 material. A more comprehensive list of examples that may be subject to industry codes and standards is at section 138 of the Act.

The material dealt with by the Expectations and industry codes of practice

The Expectations cover a broad range of online content. The Expectations can be applied to the range of material or activity on a service that is or may be unlawful or harmful. Section 46 of the Act sets out some examples of the type of material Government considers particularly harmful to Australian users, and fall within scope of the Expectations. These are:

- Cyberbullying material targeted at an Australian child;
- Cyber-Abuse material targeting an Australian adult;
- Image-based abuse material;
- Class 1 and Class 2 content; and
- Material depicting, promoting, inciting or instructing in abhorrent violent conduct.

Industry codes have a narrower scope than the Expectations. The industry codes and standards are focused on class 1 and class 2 material as defined in Part 9 of the Act. Section 138 emphasises end-user empowerment, including education, the provision of information and filtering methods as some of the most practical means by which responsible adults can facilitate appropriate controls, particularly in the case of children.

Who is subject to the Expectations and the industry codes of practice?

The Expectations and the codes apply to industry differently. The Expectations are limited to providers of social media services, relevant electronic services and designated internet services.

In contrast, section 135 of the Act lists the sections of the online industry whose peak bodies or associations are expected to develop codes or comply with a mandatory standard. These include providers of social media services, relevant electronic services, designated internet services, internet search engine services, app distribution services or hosting services, providers of internet carriage services and those that manufacture, supply, maintain or install equipment that is for use by end-users in Australia of a social media service, relevant electronic service, designated internet service or internet carriage service.

That is, all sections of the online services industry will be required to do their part to protect users from exposure to content that may harm them.

Consultation

The draft Determination is available on the Department of Infrastructure, Transport, Regional Development and Communications website. The Department invites submissions on the draft Determination from stakeholders and other interested parties by 5pm on Friday 15 October 2021. Submissions may be lodged in the following ways:

Website www.communications.gov.au/have-your-say/draft-online-safety-basic-online-safety-expectations-determination-2021-consultation

Email OnlineSafety@infrastructure.gov.au

Post Director, Online Safety Reform and Research Section
Department of Infrastructure, Transport, Regional Development and Communications
GPO Box 2154
Canberra ACT 2601

Submissions should include your name, organisation (if relevant) and contact details. The Department will not consider submissions without verifiable contact details.

Submissions will be treated as non-confidential information, and will be made publicly available on the Department's website unless you specifically request that your submission, or part of a submission, be kept confidential, and provide acceptable reasons. An email disclaimer asserting confidentiality of the entire submission is not sufficient, nor is a header or footer disclaimer.

The Department reserves the right not to publish a submission, or any part of a submission, at its absolute discretion. The Department will not enter into any correspondence with respondents in relation to any decisions not to publish a submission in whole or in part.

The Department is subject to the *Freedom of Information Act 1982* and may be required to disclose submissions in response to requests made under that Act.

The *Privacy Act 1988* establishes certain principles regarding the collection, use and disclosure of information about individuals. Any personal information respondents provide to the Department through submissions will be used for purposes related to considering issues raised in this paper, in accordance with the Privacy Act. If the Department makes a submission, or part of a submission, publicly available, the name of the respondent will be included. Respondents should clearly indicate in their submissions if they do not wish their name to be included in any publication relating to the consultation that the Department may publish.

Questions about the submission process can be directed to OnlineSafety@infrastructure.gov.au