



Online Safety (Basic Online Safety Expectations) Amendment Determination 2023

I, Michelle Rowland, Minister for Communications, make the following determination.

Dated

Michelle Rowland **DRAFT ONLY—NOT FOR SIGNATURE**
Minister for Communications

Contents

1 Name.....	1
2 Commencement.....	1
3 Authority.....	1
4 Schedules.....	1
Schedule 1—Amendments	2
<i>Online Safety (Basic Online Safety Expectations) Determination 2022</i>	2

1 Name

This instrument is the *Online Safety (Basic Online Safety Expectations) Amendment Determination 2023*.

2 Commencement

This instrument commences the day after this instrument is registered.

3 Authority

This instrument is made under section 45 of the *Online Safety Act 2021*.

4 Schedules

Each instrument that is specified in a Schedule to this instrument is amended or repealed as set out in the applicable items in the Schedule concerned, and any other item in a Schedule to this instrument has effect according to its terms.

Schedule 1—Amendments

Online Safety (Basic Online Safety Expectations) Determination 2022

1 After subsection 6(2)

Insert:

Additional expectation

- (2A) The provider of the service will take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any service that is used by, or accessible to, children.

2 Subsection 6(3)

Omit “without limiting subsection (1) or (2), reasonable steps for the purposes of this section” and substitute “without limiting subsection (1), (2) or (2A), reasonable steps for the purposes of those subsections”.

3 Paragraph 6(3)(e)

Repeal the paragraph, substitute:

- (e) ensuring that assessments of safety risks and impacts are undertaken, identified risks are appropriately mitigated, and safety review processes are implemented, throughout the design, development, deployment and post-deployment stages for the service;

4 After paragraph 6(3)(e)

Insert:

- (f) assessing whether business decisions will have a significant adverse impact on the ability of end-users to use the service in a safe manner and in such circumstances, appropriately mitigating the impact;
- (g) having staff, systems, tools and processes to action reports and complaints within a reasonable period of time in accordance with subsection 14(3);
- (h) investing in systems, tools and processes to improve the prevention and detection of material or activity on the service that is unlawful or harmful;
- (i) having processes for detecting and addressing hate speech which breaches a service’s terms of use and, where applicable, breaches a service’s policies and procedures and standards of conduct mentioned in section 14.
- (4) For the purposes of paragraph 6(3)(i), hate speech is communication by an end-user that breaches a service’s terms of use and, where applicable, breaches a service’s policies and procedures or standards of conduct mentioned in section 14, and can include communication which expresses hate against a person or group of people on the basis of race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, gender identity, disease, immigrant status, asylum seeker or refugee status, or age.

Additional expectation

- (5) The provider of the service will take reasonable steps to make available controls that give end-users the choice and autonomy to support safe online interactions.

Examples of reasonable steps that could be taken

- (6) Without limiting subsection (5), reasonable steps for the purposes of that subsection include the following:
 - (a) making available blocking and muting controls for end-users;
 - (b) making available opt-in and opt-out measures regarding the types of content that end-users can receive;
 - (c) enabling end-users to make changes to their privacy and safety settings.

5 After section 8

Insert:

8A Additional expectations—provider will take reasonable steps regarding generative artificial intelligence capabilities

- (1) If the service uses or enables the use of generative artificial intelligence capabilities, the provider of the service will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of artificial intelligence capabilities on the service.
- (2) If the service uses or enables the use of generative artificial intelligence capabilities, the provider of the service will take reasonable steps to proactively minimise the extent to which generative artificial intelligence capabilities may be used to produce material or facilitate activity that is unlawful or harmful.

Examples of reasonable steps that could be taken

- (3) Without limiting subsection (1) or (2), reasonable steps for the purposes of this section could include the following:
 - (a) ensuring that assessments of safety risks and impacts are undertaken, identified risks are appropriately mitigated, and safety review processes are implemented throughout the design, development, deployment and post-deployment stages of generative artificial intelligence capabilities;
 - (b) providing educational or explanatory tools (including when new features are integrated) to end-users that promote understanding of generative artificial intelligence capabilities on the service and any risks associated with the capabilities;
 - (c) ensuring that training materials for generative artificial intelligence capabilities and models do not contain unlawful or harmful material;
 - (d) ensuring that generative artificial intelligence capabilities can detect and prevent prompts that generate unlawful or harmful material.

8B Additional expectations—provider will take reasonable steps regarding recommender systems

- (1) If the service uses recommender systems, the provider of the service will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of recommender systems on the service.
- (2) If the service uses recommender systems, the provider of the service will take reasonable steps to ensure that recommender systems are designed to minimise the amplification of material or activity on the service that is unlawful or harmful.

Examples of reasonable steps that could be taken

- (3) Without limiting subsection (1) or (2), reasonable steps for the purposes of this section could include the following:
- (a) ensuring that assessments of safety risks and impacts are undertaken, identified risks are appropriately mitigated, and safety review processes are implemented throughout the design, development, deployment and post-deployment stages of recommender systems;
 - (b) providing educational or explanatory tools (including when new features are integrated) to end-users that promote understanding of recommender systems on the service, their objectives, and any risks associated with such systems;
 - (c) enabling end-users to make complaints or enquiries about the role recommender systems may play in presenting material or activity on the service that is unlawful or harmful.

6 Paragraph 9(2)(a)

Repeal the paragraph, substitute:

- (a) having processes, including proactive processes, that prevent the same person from repeatedly using anonymous accounts to post material, or to engage in activity, that is unlawful or harmful;

7 Subsection 10(1)

Repeal the subsection, substitute:

- (1) The provider of the service will take reasonable steps to consult and cooperate with providers of other services, and to take reasonable steps to ensure consultation and cooperation occurs between all relevant services provided by that provider in order to promote the ability of end-users to use all of those services in a safe manner.

8 Paragraphs 10(2)(a) and (b)

Repeal the paragraphs, substitute:

- (a) working with other service providers and between all relevant services provided by a service provider to detect high volume, cross-platform attacks (also known as volumetric or ‘pile-on’ attacks);
- (b) sharing information with other service providers and between all relevant services provided by a service provider on material or activity on the service that is unlawful or harmful, for the purpose of preventing and dealing with such material or activity.

9 Paragraph 12(2)(a)

Repeal the paragraph, substitute:

- (a) implementing appropriate age assurance mechanisms;

10 After paragraph 12(2)(b)

Insert:

- (c) continually seeking to develop, support or source, and implement improved technologies and processes for preventing access by children to class 2 material.

11 After subsection 14(1)

Insert:

- (1A) The provider of the service will take reasonable steps (including proactive steps) to detect breaches of its terms of use and, where applicable, breaches of policies and procedures in relation to the safety of end-users, and standards of conduct for end-users.

12 Subsection 14(2)

Repeal the subsection, substitute:

- (2) The provider of the service will take reasonable steps (including proactive steps) to ensure that any penalties specified for breaches of its terms of use, policies and procedures in relation to the safety of end-users, and standards of conduct for end-users, are enforced against all accounts held or created by the end-user who breached the terms of use and, where applicable, breached the policies and procedures, and standards of conduct, of the service.

13 After subsection 14(2)

Insert:

- (3) The provider of the service will, within a reasonable period of time:
 - (a) review and respond to reports and complaints mentioned in sections 13 and 15; and
 - (b) provide feedback on the action taken.
- (4) For the purposes of subsection (3), in determining ‘a reasonable period of time’, the provider must have regard to:
 - (a) the nature and impact of the harm that is the subject of the report or complaint;
 - (b) the complexity of investigating the report or complaint; and
 - (c) any other relevant matters.
- (5) For the purposes of paragraph (3)(a):
 - (a) *review* means considering a report or complaint from when it is first made; and
 - (b) *respond* means taking and implementing a decision to have content removed and reported, have an end-user banned, or other content moderation decisions, or a decision to take no action.

14 Subsection 15(2)

Repeal the subsection, substitute:

- (2) The provider of the service will ensure that the service has clear and readily identifiable mechanisms that enable any person ordinarily resident in Australia to report, and make complaints about, breaches of the service’s terms of use and, where applicable, breaches of the service’s policies and procedures and standards of conduct mentioned in section 14.

15 After section 18

Insert:

18A Additional expectation – provider will publish transparency reports

- (1) The provider of the service will publish regular transparency reports, at regular intervals of no less than 1 month and no more than 12 months, with information regarding:
 - (a) the service’s enforcement of its terms of use, policies and procedures and standards of conduct mentioned in section 14;
 - (b) the safety tools and processes deployed by the service (including in relation to a service’s key features), and their effectiveness;
 - (c) metrics on the prevalence of harms, reports and complaints, and the service’s responsiveness; and
 - (d) the number of active end-users of the service in Australia (including children) each month during the relevant reporting period.
- (2) For the purpose of paragraphs 1(a) to (d), the information and data contained in a transparency report must be specific to Australia, unless to do so is not reasonably practicable.
- (3) For the purposes of this section, a transparency report must:
 - (a) identify each relevant service a provider provides (where applicable);
 - (b) set out the information regarding the matters in paragraphs (1)(a) – (d) separately, in respect of each service provided by a provider; and
 - (c) be published within a reasonable time of the end of the relevant reporting period to which the report relates.

16 After subsection 20(4)

Additional expectation

- (5) Notwithstanding section 18A, if the Commissioner, by written notice given to a provider of the service, requests the provider to give the Commissioner a report on the number of active end-users of the service in Australia (including children) during a specified period, the provider will comply with the request within 30 days after the notice of request is given.

17 After subsection 21(1)

Insert:

Note: The provider of the service is expected to have a designated contact point regardless of whether the service has staff physically located in Australia.