Fact sheet

Draft Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024

November 2023

Contents

Overview

The eSafety Commissioner has released for consultation the draft Online Safety (Relevant Electronic Services – Class 1A and 1B Material) Industry Standard 2024, referred to in this Fact Sheet as the **Relevant Electronic Services Standard**. The information in this Fact Sheet should be read in conjunction with the Discussion Paper, to inform submissions.

Which services will need to comply with the Standard?

- The draft Relevant Electronic Services Standard will cover 'relevant electronic services' as defined in section 13A of the Online Safety Act 2021 (Cth) (**the Act**). Noting that the Act's definition captures a wide range of services, the draft Relevant Electronic Services Standard seeks to provide clarity by requiring specific measures for specific categories of services with unique risk profiles.
- These categories of relevant electronic services are outlined in Table 1 and broadly reflect those in the draft Relevant Electronic Services Code developed by industry associations.

Table 1: Relevant electronic service categories

Defined categories	
Telephony relevant electronic service	A Short Message Service (SMS) or Multimedia Messaging Service (MMS) provided over a public mobile telecommunications service.
Gaming service with limited communication functionality	A service that enables end-users to play online games with each other but only allows limited sharing of material (for example, in-game images and/or pre-selected messages).
Enterprise relevant electronic service	A service being provided to an organisation to enable people within that organisation to communicate with each other.
Pre-assessed categories	
Closed communication relevant electronic service	A service that enables an end-user to communicate with another end-user, but only if they already have each other's contact details (for example, their phone number or email address). This is a broad category that includes email services, some online messaging services and some video conferencing services, as well as some carriage services (email but not text messaging).

Dating service	A service primarily used for dating that has a messaging function. This category does not include escort or sex work services
Gaming service with communication functionality	A service that enables end-users to play online games with each other and share material with each other (for example, URLs, hyperlinks, images and/or videos).
Open communication relevant electronic service	A service that enables an end-user to communicate with another user and view, navigate or search for other users without already having their contact details. This category mainly includes online messaging services and video conferencing services.

- If a relevant electronic service does not meet the criteria for any of the above categories, the service will need to undertake a risk assessment and could be classified as one of the following:
    - Tier 1 relevant electronic service: high risk
    - Tier 2 RES: medium risk
    - Tier 3 RES: low risk
- This tiered approach provides flexibility to cover future relevant electronic services which may not fall within one of the specified categories.

What material is covered by the draft Standard?

- The draft Relevant Electronic Services Standard puts in place minimum compliance measures to address, minimise and prevent harms associated with access and exposure to the most harmful forms of online material. It covers:
    - class 1A material, which comprises child sexual exploitation material, pro-terror material, and extreme crime and violence material
    - class 1B material, which comprises crime and violence material and drug-related material.[1]
- These types of material are subcategories of class 1 material under the Online Safety Act, which is material that has been or would be refused classification (**RC**) under the Classification Act. Serious harms are associated with this material whenever it is produced, distributed or consumed.

[1] Importantly, the nature of the material, including its literary, artistic or educational merit, and whether it serves a medical, legal, social or scientific purpose, is relevant to the assessment of class 1B material. Material only falls within class 1B if there is no justification for the material.

- A future industry code or industry standard will be developed to address class 2 material under the Act, such as online pornography.

How is the draft Standard different to the draft Code?

- Many of the draft Relevant Electronic Services Standard provisions will look familiar to those involved in industry development of the draft Relevant Electronic Services Code which the eSafety Commissioner declined to register (referred to in this Fact Sheet as the **draft Code**) – including parts of the Head Terms[2], the overall approach to relevant electronic service categories and risk tiers, various definitions and minimum compliance measures.
- In creating the draft Relevant Electronic Services Standard, eSafety sought to build on the extensive work of industry in developing and consulting on the draft Code. This means that where appropriate, eSafety has used elements of the draft Code as an initial basis for the Relevant Electronic Services Standard.
- However, the draft Relevant Electronic Services Standard addresses the concerns about the draft Code that were set out in the eSafety Commissioner's Statement of Reasons on 31 May 2023, as well as additional issues identified by eSafety.
- The draft Relevant Electronic Services Standard has also been prepared in accordance with good practice for legislative instruments, as well as relevant requirements for effective regulation. This means it does not have the same wording and format as the industry's draft Code. For example, detailed guidance and examples will be contained in the explanatory statement to the Relevant Electronic Services Standard and the regulatory guidance, instead of in the Standard itself.
- Under the draft Relevant Electronic Services Standard, all closed communication and open communication relevant electronic services will need to:

[2] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms In force – latest version. https://www.esafety.gov.au/sites/default/files/2023-09/Consolidated-Industry-Codes-of-Practice-Head-Terms-12-September-23.pdf.

  - use systems, processes and technologies to detect and remove 'known'[3] child sexual abuse material and 'known' pro-terror material, where technically feasible
  - use systems, processes and technologies to disrupt and deter child sexual abuse material and pro-terror material, including 'known' material and 'new'[4] material
  - have trust and safety personnel to oversee safety on their service
  - adopt appropriate safety features and provide information about these – for example, complaints reporting mechanisms for end-users
  - take steps to enforce their own policies relating to class 1A and 1B material.

- Unlike the draft Code, the draft Relevant Electronic Services Standard does not have a separate category for end-to-end encrypted services. eSafety recognises that such services may face technical limitations in detecting known child sexual abuse material and known pro-terror material. Therefore, this requirement only applies if it is technically feasible for the service to detect and remove the material. eSafety considers this is a more appropriate and flexible way of dealing with these limitations rather than having a separate category for end-to-end-encrypted services.
- Under the draft Relevant Electronic Services Standard, matters to consider in relation to technical feasibility include whether it is reasonable for a service provider to incur the costs of taking action, having regard to the level of risk to the online safety of end-users.

How does the draft Standard differentiate services based on risk?

- Consistent with the registered codes for other sections of the online industry, the draft Relevant Electronic Services Standard adopts an outcomes- and risk-based

[3] 'Known' material refers to child sexual exploitation material or pro-terror material that has been previously verified, for example by a recognised child protection organisation or by an organisation with expertise in counter-terrorism.

[4] 'New' material refers to newly produced or recorded child sexual exploitation material or pro-terror material which has not been previously verified.

approach. The measures contained in the Standard are proportionate to the risk that a service presents in respect of class 1A and 1B material.

- Similar to the draft Code, the draft Relevant Electronic Services Standard proposes compliance measures based on the risk profile of each category of relevant electronic service.
- Most relevant electronic services will fall within a category as set out in Table 1. They are deemed to have a specified risk profile, and are therefore not required to conduct a risk assessment unless they make a material change to their service.
- The risk assessment requirement for the remaining services is designed to future proof the Relevant Electronic Services Standard Standard by providing a process for new relevant electronic services that do not fall within a category as set out in Table 1 to assess their level of risk. These relevant electronic services will be required to conduct a risk assessment, unless they determine their risk profile is Tier 1.

What about privacy? Will service providers be required to monitor the content of private communications?

- The draft Relevant Electronic Services Standard does not require service providers to monitor the content of private emails, instant messages, SMS, MMS, online chats and other private communications.
- However, eSafety does require service providers to use systems, processes and technologies to detect known child sexual abuse material and known pro-terror material.
- There are multiple tools and processes that relevant electronic services can use to meet their requirements under the draft Relevant Electronic Services Standard to detect and identify harmful material, that do not require companies to view the specific content of communications. For example, hash matching technologies involve using algorithms to create a digital fingerprint of a file such as an image or video. This digital fingerprint, or 'hash', is compared against hashes of known material (including previously verified class 1A material) to find copies of the same image or video.[5] One common hash matching technology called PhotoDNA has a

[5] Dr Hany Farid 2021, An Overview of Perceptual Hashing, Journal of Online Trust and Safety Vol 1, October 2021.

false positive rate of 1 in 10 billion (this is how often the systems will flag content as child sexual abuse material when it is in fact not).[6] Many online services, including some email and private messaging services, already use such tools. eSafety recognises the importance of private communication and considers that hash matching can protect the privacy both of end-users and of child victims, whose privacy is repeatedly infringed when child sexual abuse material is shared online.

- eSafety recognises that not all services will be able to use systems, processes and technologies to detect and remove known child sexual abuse material and known pro-terror material. Where it is technically infeasible for specific relevant electronic service categories to deploy tools to automatically detect and remove known child sexual abuse material and known pro-terror material, the provider is required to take appropriate alternative action. At eSafety's request, the provider must specify where it is technically infeasible to comply, and the appropriate alternative actions applicable. Importantly, service providers must also disrupt and deter both known and new child sexual abuse material and pro-terror material.
- eSafety considers that privacy and safety are not mutually exclusive, and can both be maintained through good design.

Will the proactive detection requirements weaken end-to-end-encrypted services?

- eSafety **does not** expect companies to design systemic vulnerabilities or weaknesses into end-to-end-encrypted services.
- The draft Relevant Electronic Services Standard does not require service providers to do anything that is not technically feasible.
- The draft Code contained a blanket exemption on end-to-end-encrypted services meeting requirements to detect and remove known child sexual abuse material and known pro-terror material. eSafety has not retained this in the draft Relevant Electronic Services Standard due to the breadth of its application.
- However, eSafety recognises that end-to-end-encrypted services face technical limitations, and the draft Relevant Electronic Services Standard proposes an exclusion on the requirement to detect and remove known child sexual abuse material and known pro-terror material: this requirement (sections 21 and 22) only

[6] International Telecommunications Union, Case Study, PhotoDNA. https://www.itu.int/en/cop/case-studies/Documents/ICMEC_PhotoDNA.PDF.

applies if it is technically feasible for the service to detect and remove the material. eSafety considers this is a more appropriate and flexible way of dealing with these technical limitations rather than having a separate category for end-to-end-encrypted services.

- Further, as detection technologies are developed and tested, relevant electronic services currently unable to meet the requirements in sections 21 and 22 (including because the service is end-to-end-encrypted) may find that detection becomes feasible.
- For more information on end-to-end-encryption, see eSafety's End-to-end encryption – position statement, updated 17 October 2023.

If it's not technically feasible for a service provider to detect and remove harmful material, what requirements will it be expected to meet?

- Where it is not technically feasible for a service provider to detect and remove known child sexual abuse material and known pro-terror material, the service provider will be required to:
    - demonstrate, on eSafety's request, why automatic detection is technically infeasible in the circumstances
    - take appropriate alternative action.
- The draft Relevant Electronic Services Standard is technology-neutral and outcomes-based and does not specify particular actions and technologies to be deployed.
- Where it is not technically feasible for a service to put in place systems, processes and technologies to detect and remove known child sexual abuse material and known pro-terror material, appropriate alternative action must be taken. An example of this is end-to-end encrypted services using hash matching, machine learning, artificial intelligence and other detection technologies on parts of the service that are not encrypted (such as content in end-user reports and usernames).
- Importantly, service providers are required to disrupt and deter both known and new child sexual abuse material and pro-terror material. Examples of this include:

  - deploying safety tools that disrupt or deter the distribution of child sexual abuse material and pro-terror material
  - interventions that are targeted at preventing end-users from making this material available on the service, for example by acquiring and using off-platform information that can help identify and block the registration of potential end-users who have distributed child sexual abuse material and/or pro-terror material in other environments – this could mean providers taking into account credible information published, provided or validated by another service about significant threats posed by an end-user in relation to child sexual exploitation and abuse or terrorism.

- These, and other actions to deter and disrupt the distribution of child sexual abuse material and pro-terror material, can make a significant contribution to addressing the harms associated with this material, especially in circumstances where service providers may be limited in their ability to detect and remove known material.
- Further, relevant electronic service providers with at least 1 million monthly active users in Australia will be required to have a program of investment and development to disrupt and deter the distribution of child sexual abuse material and pro-terror material, including new material.

Will a service provider be required to comply with multiple standards and/or codes?

- Consistent with the principle in the Head Terms,[7] no service provider will have to comply with more than one industry code or one industry standard in relation to the same electronic service. This is reflected in section 5 of the draft Relevant Electronic Services Standard.
- Providers of multiple online services will be subject to the industry code or industry standard applicable for each service.

[7] Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms In force – latest version. Page 4.https://www.esafety.gov.au/sites/default/files/2023-09/Consolidated-Industry-Codes-of-Practice-Head-Terms-12-September-23.pdf

- Where a single online service could fall within the scope of more than one industry code or industry standard, the code or standard that will apply is the code or standard that the service's predominant functionality is most closely aligned with.

What happens if a service provider doesn't comply with the Standard?

- The draft Relevant Electronic Services Standard sets out minimum compliance measures which will be enforceable and backed by civil penalties, enforceable undertakings and injunctions.
- If a relevant electronic service fails to comply with the Relevant Electronic Services Standard, then eSafety may make use of its enforcement powers under the Act. Unlike the Codes for other sections of the industry, under the Relevant Electronic Services Standard eSafety can take enforcement action without first directing the provider to comply with a requirement.
- eSafety will take a graduated and proportionate approach to enforcement. eSafety's approach to enforcement will be set out in its regulatory guidance for the Relevant Electronic Services Standard.
- eSafety will be able to receive complaints and investigate potential breaches of the Relevant Electronic Services Standard. When assessing whether adopted compliance measures are reasonable, eSafety will consider a range of factors including the capability and size of a service provider.

When will the Standard come into effect?

- After the public consultation closes, eSafety will carefully consider all submissions and, where appropriate, amend the draft Relevant Electronic Services Standard.
- Depending on the public consultation, eSafety expects the Relevant Electronic Services Standard will be finalised and registered on the Federal Register of Legislation in April 2024.
- eSafety currently proposes that the Relevant Electronic Services Standard will commence six months from the time it is registered, to allow time for service providers to prepare for implementation and for eSafety to provide regulatory guidance.

How do the Industry Codes and Standards fit with the Basic Online Safety Expectations?

- The Industry Codes and Standards will impose enforceable obligations on eight sections of the online industry in relation to class 1A and class 1B material (and, in future, Class 2 material). By contrast, the Basic Online Safety Expectations provide a benchmark for preventing a broader range of online harms, setting out the Australian Government's expectations for three specific sections of the online industry: relevant electronic services, designated internet services and social media services.
- Relevant electronic services are covered by both the Industry Standards and the Basic Online Safety Expectations, which are designed to complement each other. Compliance with the requirements of the Relevant Electronic Services Standard will be pertinent to a service provider's implementation of some of the Expectations but will not determine whether it meets the Basic Online Safety Expectations.

eSafety.gov.au