Summary of Reasons – Social Media Services Code

31 May 2023

eSafety decision

The eSafety Commissioner (**eSafety**) has decided to register the *Social Media Services Online Safety Code (Class 1A and Class 1B Material)* (**SMS Code**). The SMS Code meets the statutory requirements set out in section 140 of the *Online Safety Act 2021* (Cth) (the **Act**).

Background

The Act permits eSafety to register an industry code that has been developed and submitted by a body or association that represents a particular section of the online industry. To register an industry code, eSafety must be satisfied that it meets the requirements under section 140 of the Act, including that it provides appropriate community safeguards for any matters of substantial relevance to the community.

On 11 April 2022, eSafety gave a notice to the Communications Alliance and Digital Industry Group Inc (the **Applicants**) under section 141 of the Act requesting that they develop an industry code dealing with certain matters (the **Notice**).

On 18 November 2022, the Applicants submitted a draft of the SMS Code to eSafety pursuant to the Notice. In February 2023, eSafety gave a statement of preliminary views on this draft to the Applicants and invited the Applicants to submit a final version addressing feedback in eSafety's statement.

On 31 March 2023, the Applicants submitted the SMS Code to eSafety for registration, with a covering document entitled 'Request for Registration of Online Safety Codes' (the **Request**).

Scope of the SMS Code

The SMS Code applies to providers of social media services (**SMS Providers**). Social media services[1] (**SMS**) are services where the sole or primary purpose is to enable online social interaction and where the service allows end-users to interact with other end-users and post material.

The SMS Code contains measures to address, minimise and prevent harms associated with access and exposure to the most harmful forms of online material on social media services. Material covered by the SMS Code includes:

[1] The definition of SMS in the SMS Code matches the definition of social media service in section 13 of the Act.

- **class 1A material**, which is comprised of child sexual exploitation material, pro-terror material, and extreme crime and violence material, and
- **class 1B material**, which is comprised of crime and violence material and drug-related material,

in each case as described in Annexure A to the SMS Code Head Terms, which reflects the *Classification (Publications, Films and Computer Games) Act 1995* (Cth) (**Classification Act**) and related instruments.[2]

These types of material are subcategories of class 1 material under the Act, which is material that has been or would be refused classification under the Classification Act. Serious harms are associated with these kinds of material whenever it is produced, distributed or consumed.

A future industry code or industry standard will be developed to address class 2 material under the Act, which includes material that has been or would be classified X 18+, R 18+, Category 1 Restricted or Category 2 Restricted under the Classification Act.

Structure of the SMS Code

The SMS Code provides different compliance measures based on different services' risk profiles. SMS Providers are required to self-assess the risk to Australian end-users[3] that class 1A and 1B material will be accessed, distributed, or stored on their services. In making this assessment, SMS Providers must consider a range of factors, including:

- the need to be objective in evaluating the risk of harm
- a forward-looking analysis of their operating environment
- the geographical spread of their users, the age of the user base, and
- relevant international laws and guidance.

The assessment is to be carried out by responsible persons with the right level of skills, experience, and expertise.

Three tiers of risk are identified in the SMS Code. Broadly speaking, where a service:

- has over 3 million Australian end-users and over 30 million users globally[4]
- is intended for general social interaction
- enables sharing of materials in all formats, and
- enables end-users to navigate to other end-users' connections,

it is *expected* that the service will be a **Tier 1** SMS. Tier 1 services have the most requirements and are the only tier required to take proactive steps to detect known (i.e. pre-identified) child sexual abuse and known pro-terror material.

[2] 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 – see section 11 of the Classification Act. Material only falls within class 1B if there is no justification for the material.

[3] 'Australian end-user' is used throughout the industry codes but is defined in clause 2 of the Head Terms as an end-user in Australia to align with the language and scope of the Act. Both terms are used in this document.

[4] The number of Australian and global end-users are to be identified according to the number of monthly active account holders.

Services with a user base of between 500,000 and 3 million Australian end-users (and 5-30 million users globally) and which enable sharing of materials in a variety of formats, excluding live video streaming, are broadly expected to be **Tier 2** SMS.[5]

Certain services will be deemed **Tier 3** SMS where:

- an integrated chat or messaging service is not provided
- the primary purpose is for social interaction within a limited user group (or within a commercial or public enterprise), and
- the service does not enable and end-user to:
    - create a list of other end-users with whom they share a connection
    - view and navigate to a list of another end-user's connections, or
    - construct a public or semi-public profile.

Other services may also be Tier 3 where:

- the primary purpose is for social interaction within a limited user group (or within a commercial or public enterprise),
- there are less than 500,000 Australian end-users (and less than 5 million worldwide)
- the sharing of images or videos is not enabled, and
- the service does not enable an end-user to carry out the specific actions identified above (e.g. create a list of other end users with whom they have a connection).

Services falling within Tier 3 are not required to conduct a risk assessment and will be subject to minimal requirements under the SMS Code.

eSafety assessment of the SMS Code

SMS Providers, particularly those providing popular services, can perform a critical role in reducing the accessibility and dissemination of class 1A and class 1B materials.

The SMS Code sets out a range of minimum compliance measures for SMS Providers that the Applicants submit provide appropriate community safeguards for the matters identified in the Request.

eSafety agrees that the matters identified by the Applicants in the Request are matters of substantial relevance to the community, and that the SMS Code provides appropriate community safeguards for those matters.

In particular, eSafety agrees that a tiered approach to the obligations on SMS is proportionate and appropriate. Larger and higher-risk SMS with a broad user base and which have features that increase risk (e.g. integrated chat functions), such as Facebook, Instagram, Twitter and TikTok, should have more obligations than small community-based SMS, particularly those focused on a common singular interest.

[5] The SMS Code has a list of factors which are to be used by SMS Providers as a guide for developing a risk assessment methodology, not all of which are covered here. See clause 5 of the SMS Schedule for a complete list.

While there is some uncertainty with the application of the risk criteria given that services are required to self-assess and there are multiple criteria identified (none of which is determinative), the SMS Code provides that, if the risk assessment indicates that a service is between risk tiers, the provider must apply the higher risk profile (tier) to that service. eSafety is also able to require Tier 2 SMS to provide details of their risk assessment. On balance, eSafety considers that the flexibility inherent in this risk assessment is appropriate given the broad nature of the services that may be considered an SMS.

eSafety has found that the SMS Code provides appropriate community safeguards in relation to each of the matters identified by the Applicants for the following reasons:

1. SMS are required to take reasonable and proactive steps to create and maintain a safe online environment for Australian end-users
2. SMS are required to empower their users to manage access and exposure to class 1A and class 1B material it, and
3. the SMS Code strengthens transparency of, and accountability of SMS for, class 1A and class 1B material.

eSafety's assessment has focused on the level of requirements for each tier of SMS Providers.

Obligations on Tier 1 and Tier 2 SMS Providers

The obligations on Tier 1 and Tier 2 SMS Providers include requirements to:

- have systems, processes and technologies that enable the provider to take enforcement action against end-users who breach terms and conditions or policies that prohibit class 1A and class 1B material. The provider must also take appropriate enforcement measures against end-users when this occurs
- remove child sexual exploitation material and pro-terror material within 24 hours of it being identified and take enforcement action against those distributing such material, including terminating accounts and preventing the creation of further accounts
- implement minimum Safety by Design[6] features and settings, including settings that are designed to prevent unwanted contact from other end-users (in the case of a Tier 1 SMS, these are to be the default settings for Australian children)
- enforce policies regarding minimum age permitted on the service and ensure only registered account holders can post on the service
- provide clear and accessible online safety information which include:
    - the safety management tools and settings available to parents (where the services permit children to hold accounts)
    - the role and function of eSafety and how to make a complaint and seek support
    - how to make a report regarding class 1A or 1B material, and
    - how to make a complaint regarding compliance with the SMS Code
- provide tools which enable Australian end-users to report or make a complaint about

[6] Safety by Design is an approach to development of online services and products that puts user safety and rights at the centre of the design process. For more information and relevant tools, see eSafety's Safety by design page.

class 1A or class 1B material to the service provider and also tools that enable complaints about the provider's handling of the original complaint

- promptly respond to Australian end-users in relation to reports or complaints, have appropriate policies and procedures in place and also train personnel to handle such complaints
- be reasonably resourced with personnel to oversee the safety of their service and compliance with the SMS Code
- take part in annual industry forums to share best practices, and
- contribute to expert groups that tackle child sexual exploitation and pro-terror material.

Tier 1 and Tier 2 SMS also have reporting requirements. In the case of a Tier 1 SMS, an annual code report must be submitted to eSafety detailing the steps taken to comply with their obligations under the SMS Code and the volume of child sexual exploitation material and pro-terror material removed. In the case of a Tier 2 SMS, a code report must be submitted on request from eSafety.

Tier 1-specific obligations

The obligations on Tier 1 SMS Providers are more extensive and include requirements to:

- have on-platform reporting tools enabling complaints about class 1A and 1B material on the service
- establish a central place on their service to provide online safety information to end-users in Australia
- have appropriate systems, processes and/or technology in place to proactively detect known child sexual abuse material and pro-terror material
- take action that aims to further disrupt or detect child sexual abuse material and pro-terror material (including new/first-generation child sexual abuse material) and invest in systems, processes and/or technologies that do this, and
- provide default safety and location settings to prevent end-users under the age of 16 years from unwanted contact from unknown end-users.

Obligations on all SMS Providers

The obligations on Tier 3 services are not extensive. eSafety considers this appropriate and proportionate given the breadth of services that fall within the definition of SMS. Tier 3 SMS (along with Tier 1 and 2 SMS) must:

- promptly notify law enforcement or appropriate non-government organisations of child sexual exploitation or pro-terror material on their service when there is an immediate threat to a child/adult in Australia (this is intended to supplement existing laws), and
- conduct risk assessments (e.g. using eSafety's Safety by Design tools) to ensure they continue to provide safeguards appropriate to their service.

eSafety considers that these obligations will, together, create effective and meaningful obligations on SMS in addressing the risks associated class 1A and class 1B material.

Next steps

The SMS Code will shortly be published on the eSafety Commissioner's Register of industry codes and industry standards. The SMS Code will come into effect six months after registration.