9 February 2023

John Stanton
Chief Executive Officer
Communications Alliance

Sunita Bose
Managing Director
Digital Industry Group Inc

By email: at [redacted] **and** [redacted]

Invitation to respond and/or submit amended draft code - Social Media Services

Dear John and Sunita,

On 18 November 2022, I received a request from the six industry associations making up the Steering Group (**Steering Group**) to register the Consolidated Industry Codes of Practice for the Online Safety Industry (Class 1A and Class 1B Material) purusant to section 140 of the *Online Safety Act 2021 (Cth)* (**the Act**).

As presented, the Consolidated Industry Codes comprise a set of head terms, and eight separate industry codes that apply to different sections of the online industry. In the request for registration, Communications Alliance and Digital Industry Group Inc (collectively, the **Industry Bodies**) indicated that together they represent providers of social media services and were responsible for developing the Social Media Services Online Safety Code (Class 1A and Class 1B Material) (**draft SMS Code**).

Section 140 of the Act gives me, as the eSafety Commissioner, power to register an industry code. I have considered the relevant requirements under the Act, taking into account the Steering Group's submission and accompanying documents.

I have not yet made a decision whether to register the draft SMS Code, but have formed a preliminary view. The **attached** statement sets out my preliminary views on the draft SMS Code and provides you with an opportunity to respond and/or submit an amended draft code before I finalise my decision. Separate letters will be sent to the relevant industry associations in relation to each draft industry code.

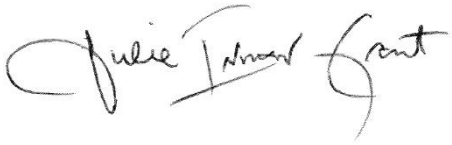
If I decide not to register a code for a section of the online industry, I intend to determine an industry standard under section 145 of the Act for that section of the online industry.

Next steps

I invite you to respond to this letter and/or submit an amended draft SMS Code by 5pm AEDT on 9 March 2023.

If you have any questions about this letter, please contact Morag Bond, Executive Manager, Legal MarComms and Research [redacted], or Maggie Law, Co-manager, Industry Codes team, on at [redacted], or the eSafety Industry Codes Team at [redacted]

Yours faithfully,



Julie Inman Grant
eSafety Commissioner

Statement of Preliminary Views – Social Media Services (SMS) Code

Summary

On the information currently available, the eSafety Commissioner's preliminary view is that:

- the draft SMS Code does not meet the requirement under s 140(1)(b) of the Act, because the code is expressed to apply in respect of 'Australian end-users' rather than 'end-users in Australia',
- the draft SMS Code does not meet the requirement under s 140(1)(d) of the Act, because it does not provide appropriate community safeguards for matters of substantial relevance to the community (as identified in the Request for registration), namely Matters 1, 4, and 11, and
- as a result, the eSafety Commissioner's jurisdiction to register an industry code under s 140(2) is not enlivened.

Industry Bodies are invited to provide a response to this Statement of Preliminary Views and submit an amended industry code addressing the areas of concern set out below.

Background

1. On 11 April 2022, the eSafety Commissioner (**eSafety**) issued a notice to the Industry Bodies, requesting the development of an industry code that applies to participants in the group consisting of providers of social media services, so far as those services are provided to end-users in Australia (as defined under s 135(2)(a)).
2. The notice required an industry code dealing with specified matters to be submitted to the eSafety Commissioner by close of business on 9 September 2022. By variation issued on 24 June 2022, eSafety extended the due date for submission of the industry code to 18 November 2022.
3. By email dated 18 November 2022, the Industry Bodies submitted the draft SMS Code to eSafety for registration. Accompanying the draft SMS Code were a cover letter, an explanatory document titled 'Request for registration', and a submission log from the public consultation and industry associations' responses to public consultation.

Section 140 requirements

4. eSafety has reviewed the Industry Bodies' submission including the accompanying documents. eSafety has also closely considered the draft SMS Code in light of previous discussions with members of the Steering Group, as well as other factors such as the current industry practice and the technological tools available and used by SMS service providers. eSafety has closed considered the effectiveness and the enforceability of the proposed compliance measures.
5. Section 140(1) of the Act sets out the conditions which must be met in order to enliven eSafety's discretionary power under s 140(2) to register a code. Based on the information currently available, the eSafety Commissioner is unlikely to be satisfied that all of the conditions in s 140(1) are met. Consequently, the power to register an industry code under s 140(2) of the Act would not be enlivened. The reasons for this are set out below.

Section 140(1)(b) requirement

6. Section 140(1)(b) of the Act requires eSafety to be satisfied that an industry code submitted by a body or association referred to in s 140(1)(a) applies to participants in that section of the online industry and deals with one or more matters relating to the online activities of those participants.
7. The relevant 'section of the online industry' for the draft SMS Code is the group consisting of providers of social media services, so far as those services are provided to end-users in Australia, as described in section 135(2)(a).[1]
8. The relevant 'online activity' for the draft SMS Code is providing a social media service, so far as the service is provided to end-users in Australia, as defined in s 134(a).
9. Clause 2.1(a) of the draft SMS Code stipulates its scope to apply to 'a provider of social media services, so far as materials on that service are provided to *Australian end-users*'. The head terms define 'Australian end-user' to mean an end-user who is ordinarily resident in Australia.
10. eSafety considers 'end-users in Australia' and 'Australian end-users' to be materially different concepts, despite the likely overlap, because the former reflects an end-user's geographical location, while the latter (as defined in the head terms) reflects the ordinary residency status of the end-user.
11. While some parts of the Act refer to 'Australians' and end-user who is 'ordinarily resident in Australia', the provisions identifying the sections of industry and online activities subject to the proposed codes (ss 134-135) are not expressed in these terms. eSafety considers that the registration criteria in s 140 must be considered by reference to ss 134-135.
12. eSafety considers it is unlikely the draft SMS Code would satisfy s 140(1)(b) of the Act because the code is expressed to apply in respect of 'Australian end-users' and not to the relevant group of providers, described in s 135(2)(a), or to the relevant online activity, described in s 134(a).

Section 140(1)(d) requirement

13. Section 140(1)(d)(i) of the Act requires eSafety to be satisfied that, to the extent to which the draft SMS Code deals with one or more matters of substantial relevance to the community, the code provides appropriate community safeguards for that matter or those matters.
14. eSafety considers the draft SMS Code will unlikely meet the requirement under s 140(1)(d)(i) of the Act, because it does not provide appropriate community safeguards for Matters 1, 4, and 11 for the reasons outlined below.

Matter 1

Measures directed towards achieving the objective of ensuring that industry participants have scalable and effective policies, procedures, systems and technologies in place to take reasonable and proactive steps to detect and prevent access or exposure to, distribution of, and online storage of class 1A material.

15. The draft SMS Code proposes minimum compliance measures (**MCMs**) 1 - 10 to deal with Matter 1.

[1] For the avoidance of doubt, eSafety is satisfied at this stage that the requirement under s 140(1)(a) that the Industry Bodies represent providers of social media services, so far as those services are provided to end-users in Australia, has been met. This Statement of Preliminary Views relates only to the scope of the draft SMS Code as submitted

16. eSafety considers that in order to provide appropriate community safeguards for Matter 1, the draft SMS Code would need to ensure, at a minimum:

    (a) all Tier 1 SMS (regardless of whether they meet the definition of Very large SMS) use systems, processes and/or technologies to detect and remove known pro-terror material/ Terrorist and Violent Extremist Content (**TVEC**); and

    (b) Tier 1 SMS make ongoing investment in systems and processes and technologies in relation to class 1A material (including first generation CSAM[2]).

17. eSafety considers that all Tier 1 SMS (regardless of whether they meet the definition of Very large SMS) could reasonably comply with a requirement to use systems, processes and/or technologies to proactively detect known TVEC/pro-terror material. While this commitment would not require providers to use all three of systems and processes and technologies, eSafety notes that in practice, some SMS providers may use the same systems, processes and technologies to detect and remove known TVEC as they do for known CSAM (while using different datasets).
18. There are multiple ways this broadly drafted requirement (which does not require the deployment of specific technology) could be met, including options appropriate for less well-resourced businesses.

    - There is a mix of proprietary and open-source tools that are widely used (we note over 200 organisations use PhotoDNA). Hash matching is considered a privacy-preserving method with minimal risk of false positives.[3]
    - SMS providers have the option to work with a recognised Non-Governmental Organisation to access hash databases. While the only current database of known TVEC accessible by multiple companies is held by the Global Internet Forum to Counter Terrorism (GIFCT), eSafety also understands Meta and Google recently introduced tools to help combat the spread of terrorist content and will make these available to a wide range of companies for free.
    - Alternatively, SMS providers could consider working with or joining Tech Against Terrorism, using the Terrorist Content Analytics Platform (as several do already) which is designed with the aim of supporting smaller tech platforms, including by sharing alerts for known TVEC URLs.
    - Providers could also make use of proprietary technologies or develop their own tools aimed at detecting, for example, material associated with proscribed terrorist organisations based on Australian law or international reference points such as the UNSC Consolidated Sanctions List, including ensuring that content reported to their platform is not re-uploaded.

19. At present, eSafety is likely to conclude that it is not reasonable to curtail the ongoing investment requirement in MCM 10 to *known CSAM,* due to the immensely harmful consequences associated with the creation, distribution and dissemination of first generation CSAM. It is also unclear why the ongoing investment requirement is limited to *instances of videos and images that depict and promote*

[2] Child Sexual Abuse Material
[3] Testimony of Dr Hany Farid to House Committee on Energy and Commerce Fostering a Healthier Internet to Protect Consumers, 2019: www.congress.gov/116/meeting/house/110075/witnesses/HHRG-116-IF16-Wstate-FaridH-20191016.pdf

a terrorist act. Technologies and processes which detect first generation CSAM and TVEC are increasingly being developed and deployed on services. eSafety considers that commitments by Tier 1 providers to invest in the development and/or deployment of such technology is critical in order to provide appropriate community safeguards. eSafety also considers it important for the commitment to cover investment in systems, processes and technologies, given processes and systems (potentially incorporating human moderation) will need to sit alongside technologies. This approach increases the effectiveness and enforceability of this commitment.

20. eSafety notes that the commitments in the draft SMS Code also fall short of the Voluntary Principles to Counter Online Child Sexual Exploitation and Abuse (the **Principles**),[4] which reflect collective resolve and have been endorsed by some of the leading technology companies since 2020 including Snap Inc, TikTok, Yubo (which are likely to be Tier 1 but may not be a Very large SMS under the draft SMS code) as well as Meta and YouTube. The Principles were developed in conjunction with industry representatives, and in consultation with a broad range of experts from civil society and academia as part of a global response, and designed to apply across different services. The Principles are supported by the Five Country governments (Australia, Canada, New Zealand, UK and US) and the G7. In particular, the Principles endorsed by the companies include identifying and combating the dissemination of **new** child sexual abuse material via their platforms and services, and to consider where existing measures can go further and to invest in innovative tools and solutions (e.g. Principle 2).

Matter 4

Measures directed towards achieving the objective of ensuring that industry participants have scalable and effective policies, procedures, systems and technologies in place to take reasonable and proactive steps to limit the hosting of class 1A material and class 1B material in Australia.

21. The draft SMS Code does not include specific measures to deal with Matter 4. Industry Bodies submit that the other measures in the draft SMS Code that are designed to limit online storage of class 1A material by a social media service address the first party hosting of this material by such services.

22. Having regard to the measures proposed for Matter 1 and Matter 2 and the concerns set out above, our preliminary view is that in order to provide appropriate community safeguards for Matter 4, the draft SMS Code would need to ensure it contains, at a minimum, the steps in paragraph 16 above.

Matter 11

Measures directed towards achieving the objective of ensuring that industry participants publish annual reports about class 1A material and class 1B material, and their compliance with industry codes.

23. The draft SMS Code proposes MCMs 32 and 33 to deal with this matter, with tier 1 SMS providers subject to more specific reporting requirements (in MCM 32) than tier 2 SMS providers (MCM 33).
24. Under MCM 33, Tier 2 SMS providers must submit a code compliance report within 6 months of

[4] Principles, including signatories retrieved at: www.weprotect.org/library/voluntary-principles-to-counter-online-child-sexual-exploitation-and-abuse/

receiving a request from eSafety, although any request that would otherwise be due within the first 12 months after the code comes into effect is not due until 12 months after the code comes into effect. The head terms further provide that a code does not come into effect until 6 months after registration. This means that no report would be due to eSafety until 18 months after registration at the earliest.

25. eSafety has concerns that the timeframe for responding to requests for reports under MCM 33 will impact eSafety's ability to consider a service provider's compliance with code commitments, as well as eSafety's ability to provide constructive input into the first review of the SMS Code, noting that all reporting data will be at least six months out of date. Further, without an effective review process, the capability of the SMS Code to provide appropriate community safeguards may be compromised.
26. eSafety's preliminary view is that the proposed 6 months' response timeframe in MCM 33 is likely to prevent MCM 33 from providing appropriate community safeguards in relation to this matter, and suggests that a response timeframe of 2 months would be appropriate.

Enforceability of the code

27. In order to provide appropriate community safeguards under s 140(1)(d) of the Act, the head terms and the specific provisions in each industry code, when read as a whole, must be capable of being implemented and being enforced. This means ensuring service providers, eSafety and other parties have sufficient certainty and clarity about the obligations under the codes. At the same time, eSafety recognises the importance of a balance between flexibility and ensuring compliance can be assessed and enforced.

28. eSafety has identified provisions in the head terms and draft SMS Code which are phrased and structured in ways that risk rendering the proposed compliance measures ineffective, or potentially impractical to measure and enforce. The following examples are not exhaustive:

    Limitation clause in the head terms

    - Clause 6.1 (c) limits the codes from requiring any industry participant to 'render methods of encryption or other information security measures less effective'. As previously communicated to Industry Bodies, eSafety has concerns that rendering 'other information security measures less effective' is too broad and is a very low bar. There is a risk that as drafted, clause 6.1(c) could create broad exclusions from code commitments. eSafety considers it important that service providers consider how code compliance could be achieved by alternative mechanisms or by remedying the design.
    - Clause 6.1 (e)(iii), (h), (i) and (j) and clause 6.2 each limit the codes from requiring industry participants to take action or engage in conduct that would violate other laws. As previously communicated to Industry Bodies, eSafety considers that the blanket exclusions are not desirable and it would be more appropriate for service providers to communicate specific concerns to eSafety when a specific issue arises as to how compliance with a code requirement may breach a law and/or explore alternative approaches to meeting the minimum compliance measures of the code while still meeting other legal requirements.

Risk assessment methodology

- In relation to the risk assessment methodology in the draft SMS Code, eSafety is concerned that SMS providers may underestimate their risk level if application of the tiers and relative weighting of the factors listed in the table is left to industry participants to determine without further guidance.
- The process to identify applicable compliance measures is entirely reliant on an effective risk assessment. While SMS providers are required to demonstrate that the compliance measures they have adopted are reasonable, it would be difficult for eSafety to critically assess risk profile assigned by the SMS provider to the online activity if those risk factors are open to broad interpretation and the risk profile adopted do not accurately reflect the risk of harm.

'Appropriate steps' in MCM 26

- Under MCM 26, Tier 1 and Tier 2 SMS must take appropriate steps to promptly respond to Australian end-users regarding action taken on reports and complaints.
- eSafety recognises that the timeliness of the actions required under this measure will depend on a number of factors such as those set out in the guidance note in MCM 26. However, eSafety considers it reasonable for MCM 26 to provide greater clarity about what 'appropriate' steps entail because there is risk arising from the uncertain language where some service providers may take ineffective steps and/or respond in an unreasonable timeframe, which will undermine the effectiveness of MCM 26 to deliver appropriate community safeguards.

Next steps

29. Industry Bodies are invited to respond to the Statement of Preliminary Views and/or submit an amended code that addresses all of the following:

    (a) the scope and application of the draft SMS Code, which should align with the language of the Act where the relevant section of the online industry and relevant online activity are defined by reference to 'end-users in Australia';

    (b) a commitment by all Tier 1 SMS (regardless of whether they meet the definition of Very large SMS) to use systems, processes and /or technologies to detect and remove known pro-terror material/ TVEC;

    (c) a commitment by Tier 1 SMS to make ongoing investment in systems and processes and technologies in relation to class 1A material (including first generation CSAM);

    (d) where the draft SMS Code requires a code compliance report to be submitted within 6 months of receiving eSafety's request, the response timeframe should be revised to 2 months;

    (e) ambiguities and inconsistencies that could undermine the integrity and enforceability of the draft SMS Code should be resolved, and further guidance provided in the areas identified by eSafety in paragraph 28 above.

30. If Industry Bodies decide not to submit an amended code but wish to provide further information, the information should clearly explain how the existing MCMs will provide appropriate community safeguards despite the express concerns identified above.

31. Any submission and revised code will need to be provided to eSafety by 5pm AEDT on 9 March 2023, in order for the eSafety Commissioner to take it into account before making her final decision. For the avoidance of doubt, eSafety makes no representations that an amended code addressing the above concerns will be registered by default.