16 February 2024

Snap Inc. Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts

Thank you for the opportunity to provide a submission responding to the proposed amendments to the Basic Online Safety Expectations (BOSE) as set out through the Online Safety (Basic Online Safety Expectations) Amendment Determination 2023 (the Amendment Determination).

This submission outlines Snap's approach to trust and safety and welcomes the proposed amendments to the BOSE. While we have taken the opportunity to comment on and provide feedback for the proposed amendments, we believe our suggestions will support their effective implementation by the industry. These include recommendations for adjustments to the draft amendments. We fully recognise the need for the BOSE to adapt to the continually evolving landscape of online enforcement and reiterate our support for the proposed changes.

Introduction to Snap and Snapchat

As a brief introduction, Snap is a technology company. The company's three core products are Snapchat, a visual messaging app that enhances people's relationships with friends, family, and the world; Lens Studio, an augmented reality (AR) platform that powers AR across Snapchat and other services; and the company's AR glasses, Spectacles. We are proud to say that 414 million people globally now use Snapchat every day to express themselves and learn about the world.

Snapchat is designed very differently to traditional social media platforms: in fact Snapchat has been designed as the antidote to such platforms. Unlike traditional social media, Snapchat does not offer an open news feed where unvetted publishers or individuals have an opportunity to broadcast hate, misinformation, or violent content. Rather, Snapchat is at heart a visual messaging application, designed for private communications (either 1:1 or in limited-size groups), with the aim of encouraging users to interact creatively with their real friends, not strangers.

Our approach to safety on Snapchat

Safety by design

Safety by design, which is about prioritising the safety of our community, is at the heart of Snapchat. Product, Policy, Legal and Engineering teams are fully involved in the product and feature development lifecycle, from conception to release.

Through our safety by design framework, we have made a range of design choices to help keep Snapchatters safe:

- We **intentionally make it harder to find and talk to others on Snapchat** compared to other platforms. Snapchatters' friends' lists are private and only visible to the individual user themselves. Snapchatters also cannot receive a message from anyone whom they have not accepted as a friend on Snapchat or who is not a contact in their phone book.
- **Most content on Snapchat is designed to delete by default**: this means that default settings are such that Snaps (visual messages) sent through Chat are no longer viewable in the app shortly after they've been opened, while chat messages and Stories are, by default, only viewable for 24 hours. This further limits how widely content can be shared.
- A **Snapchat user cannot broadcast unmoderated Snaps or Stories** to the whole Snapchat community, and if they share content with just their friends, it cannot be forwarded broadly. While we do have a publicly viewable side to the app – our Spotlight and Discover platforms for news and entertainment – content there is substantially curated and moderated, respectively, ensuring harmful or illegal content is not surfaced to large numbers of people. In addition to needing to comply with our Community Guidelines, partners in Discover must adhere to our additional guidelines and terms.
- If harmful content activity takes place anywhere on Snapchat, we have effective systems and processes to act quickly. We provide **easy-to-use in-app reporting tools** so users can notify us of potential safety issues. Last year we rolled out in-chat reporting, which enables users to quickly report specific chat messages for human review.
- Our **global Trust & Safety team, including team members embedded in our Sydney office**, work 24/7 to review user reports and take appropriate action. Sanctions for violations include warning the user, removing the content, and deactivating the violating account and prohibiting the user from creating new accounts.
- We also have an **in-house global Law Enforcement Operations team, which also includes team members embedded in our Sydney office**, dedicated to reviewing and responding to law enforcement requests for data related to their investigations, and we work closely with law enforcement agencies around the world, consistent with our legal obligations.
- We are committed to transparency and openness with what is happening on Snapchat and how we respond to it. **Snap publishes bi-annual transparency reports** detailing our response to illegal and harmful content on Snapchat. Currently, Snapchat is one of the only major platforms to provide country-specific breakdowns of content reported and enforced, including a dedicated page for Australia.

Additional safeguards and resources for young people

While we prioritise creating a safe and positive experience for Snapchatters of all ages, given Snapchat's popularity among teenagers we have dedicated significant time and resources to help ensure that younger people have a safe experience when using the app.

We have done this by implementing particular safeguards and created specific resources for teen Snapchat users (aged 13-17 years), including:

- While we already intentionally make it harder to find others on Snapchat compared to other platforms, it is even harder to find Snapchatters who are not adults. Teen Snapchat users (aged 13-17) will only show up as a "suggested friend" or in search results on Snapchat **if they have several mutual friends in common** (we increased the threshold in August 2023). With the exception of parents and caregivers via our new Family Centre suite of tools, it is not possible to view another user's friends' list.
- Snapchat sends **in-app warnings to teens** if someone whom they are less likely to know, such as when they don't share multiple mutual friends, tries to add them as a friend. Snapchat also does not allow a teen to appear in search results for another user unless they have mutual connections in common or are in their respective phone contacts.
- While we already prohibit sexually explicit content on Snapchat, for Snapchatters between the ages of 13-17 we also have additional measures in place to help **prevent sexually suggestive content from appearing**, and our Spotlight content is evaluated by human moderators for being sexually suggestive before being widely distributed. Additionally, we have developed machine learning classifiers which work to identify sexually suggestive content and filter it from the experience before human intervention.
- **Location-sharing on Snap Map is off by default**, and there is no option on Snapchat to share precise location with anyone other than one's friends, or a designated sub-set of those friends. We named the default "off" location-sharing setting "Ghost Mode", with a clear accompanying ghost icon, to ensure that the setting would be understood by younger individuals and assist them in making informed choices about whether to use Snap Map, or whether to share their location and, if so, with whom.
- We recognise that young people may be reluctant to report harmful content online, and are continually looking for ways to encourage reporting. In 2021, we launched '***Safety Snapshot***', a dedicated channel on our Discover page which provides advice for users on keeping their accounts secure in a creative and visually accessible way, and is designed to appeal to young people. Last year we launched four more Safety Snapshots, covering sextortion, sending nude photos, child sex trafficking and grooming.
- In 2022 we launched **Family Centre**, a set of tools that give parents and caregivers the ability to know who their teens are friends with on Snapchat, and with whom they have been communicating over the past 7 days, without revealing the substance of their teens' conversations. Family Centre is intended to spark meaningful conversations

amongst parents and their teens about staying safe online, while at the same time providing parents or carers with the ability to report suspicious accounts to Snap for review. Family Centre also provides Content Controls that allow parents and carers to limit teens' access to "sensitive content" in the more public areas of Snapchat.

Snap response to the proposed amendments

We overall welcome and support the Government's proposals that are set out in the Amendment Determination.

We supported the creation of the BOSE in 2021 and recognise its role in establishing a common online safety benchmark for all online safety providers, which has helped to drive up standards across the online industry. The BOSE represents an effective principles-based approach to regulation that provides platforms with flexibility in how to meet the Government's expectations, rather than being prescriptive. As we outlined in our initial submission to the original 2022 Determination, the Government's recognition that service providers are best placed to identify harmful content on their services, and to choose the best way to address these in the most responsive way, is significant.

As the online environment continues to evolve and change, it is important that the BOSE follows and adapts. We agree with the Government's objective, as outlined in the Consultation Paper, "to address emerging online safety issues and gaps in the original BOSE Determination, and to improve its overall operation". It is also consistent with Snap's 'Safety by Design' approach of placing safety at the forefront of the development cycle of our products and features and we believe that we are already in alignment with the proposed changes. For example:

- 'Safety by Design' has underpinned the development of our generative AI feature Dreams.
- We already publish bi-annual transparency reports including a dedicated page for Australia.
- The eSafety Commissioner has recognised our strong response times for reports and complaints, including an industry-leading median response time of 4 minutes for user-reported CSEAM.

We have set out below some comments and feedback on the proposed Amendment Determination that we believe will better support their effective implementation by industry, including a small number of minor recommendations for how the proposed amendments could be adjusted. Notwithstanding these comments, however, it is important for us to emphasise our support for the BOSE, the proposed amendments and our recognition of the need for the BOSE to adapt to the continually evolving online environment.

Opt-in and opt-out of content types

We support the proposed additional expectation to the BOSE at sub-clause 4(5) of the Amendment Determination that "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", as well as the helpful inclusion at sub-clause 4(6) of examples of some reasonable steps. Snap provides a range of privacy and online safety controls for Snapchatters, including who can contact them, view their story or see their location, and provides additional controls for parents and caregivers, including the ability to limit their teen's ability to view sensitive content.

However, we question the inclusion through new paragraph (b) of "making available opt-in and opt-out measures regarding the types of content that end-users can receive". For the publicly viewable side of the app, Snapchatters can already opt-in and opt-out of certain categories of advertising and can report specific content for breaching our Community Guidelines or simply because they do not like it. However, we are concerned with the inclusion, even if it is only an example, of an expectation of giving end-users the ability to opt-in or opt-out of entire categories of content. We note that common categories of content include 'animal videos' or 'news', which will have no bearing to the expectation of supporting safe online interactions, would be impacted by this proposed change to the BOSE.

We believe that creating an expectation in the BOSE that platforms provide opt-ins and opt-outs for content may be inconsistent with the social benefit to end-users of platforms delivering end-users with a diversity of content, not solely content that a particular end-user is known to like or agree with. In other words, implementing opt-out features could potentially degrade the user experience. For platforms that rely on content discovery and engagement, limiting content visibility can reduce the effectiveness of recommendation algorithms, leading to a less personalised and engaging user experience.

We would further like to highlight some potentially significant implementation challenges to the proposed change. Implementing a robust 'content opt-out' mechanism will involve a high degree of technical complexity. It would require sophisticated systems, such as content tagging and AI tools, to accurately identify and filter out content based on user preferences, which can vary widely. Given the vast amount of content generated on many platforms, ensuring accurate categorisation and delivering user opt-out preferences without errors is a significant challenge.

Snap recommendation 1: We recommend that the example of making available opt-in and opt-out measures regarding the types of content that end-users can receive, as set out in paragraph 4(6)(b) of the Amendment Determination, be removed.

Appropriate age assurance mechanisms

We support the proposed clause 9 of the Amendment Determination to include the word "appropriate" in the example at subsection 12(2) of the BOSE of "implementing appropriate age assurance mechanisms". We note that the proposed amendment does not mandate that age assurance mechanisms be implemented. On this topic, we draw attention to the Government's August 2023 response to the Australian eSafety Commissioner's Roadmap for Age Verification, which noted the findings in the Roadmap that "age assurance technologies are immature, and present privacy, security, implementation and enforcement risks" and that "a decision to mandate age assurance is not ready to be taken".

Continually improving technology and practices

We support the inclusion at clause 10 of the Amendment Determination of a new example of a reasonable step at subsection 12(2) of the BOSE of "continually seeking to develop, support or source, and implement improved technologies and processes for preventing access by children to class 2 material". However, we recommend that the additional example be amended to "continually improving technology and practices" for consistency with existing language at paragraph 6(3)(d) of the BOSE. We believe this change would simplify this example without having any impact on its scope or the expectation set on industry.

Snap recommendation 2: We recommend that the proposed new paragraph 12(2)(c) of the BOSE be changed from "continually seeking to develop, support or source, and implement improved technologies and processes for preventing access by children to class 2 material" to "continually improving technology and practices for preventing access by children to class 2 material".

Business decisions

We support the inclusion at clause 4 of the Amendment Determination of a new example of a reasonable step in subsection 6(3) of the BOSE of "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". While the discussion paper makes it clear that business decisions could include "changes to terms of use to allow previously prohibited material as well as major staff cuts", which we also agreed with, we are concerned with the broad and imprecise nature of the term "business decisions" in the proposed amendment.

For example, the term "business decisions" would by definition also include a myriad of minor decisions that will have no bearing to the online safety of a provider's platforms, such as the

creation of a new sales team, relocation of an office, or a design change to a platform's logo. It would be unnecessary and unrealistic for all such decisions to be subject to this expectation.

We instead recommend replacing "business decisions" with terminology with greater specificity such as "major staffing, operational or policy decisions" to avoid regulatory burden on minor activities that have no relevance to online safety. We note that the two examples provided in the discussion paper mentioned above would clearly fall within the scope of this new wording.

Snap recommendation 3: We recommend that the proposed new paragraph 6(3)(f) of the BOSE be changed from "business decisions" to terminology with greater specificity such as "major staffing, operational or policy decisions".

Proactive processes

We note the inclusion of the term "proactive steps" in several parts of the Amendment Determination, including in clause 11 that inserts a new expectation in the BOSE for providers to take reasonable steps (including proactive steps) to detect breaches of its terms of use, policies, procedures and standards of conduct, as well as in clause 12 to insert the term "proactive steps" into the existing expectation for providers to take reasonable steps to ensure that penalties for breaches are enforced against all accounts held or created by perpetrators. While we generally support the expectation for platforms to undertake proactive steps wherever practicable, they will not always be the most appropriate or best approach in every circumstance.

For example, due to the need to protect user privacy, the use of proactive detection in private text communication may need to be more limited. Proactive monitoring often involves the automated scanning and analysis of user content, which raises significant privacy issues, particularly in relation to chat where users have an expectation of privacy. Proactive tools may also lack a sufficiently nuanced understanding of language, context, and cultural differences for accurately undertaking proactive monitoring. This can result in false positives (incorrectly flagging acceptable content as inappropriate) or false negatives (failing to identify actual violations), leading to inconsistent and unfair content moderation decisions that can harm users.

Rather, proactive detection may only be one possible way of taking action against harmful behaviour. While we support the inclusion of "proactive steps" in clauses 11 and 12, we recommend that "proactive steps" be instead listed as an example of how the above two expectations can be met, rather than incorporated into the expectations themselves.

Snap recommendation 4: We recommend that the proposed new subsection 14(1A) of the BOSE be changed so that "proactive steps" is not listed as part of the proposed new requirement but as an example of how the expectation can be met. We similarly recommend that the proposed amendment to subsection 14(2) of the BOSE also be changed so that

"proactive steps" is not listed as part of the proposed amended expectation but as an example of how the expectation can be met.

Definition of "hate speech"

We note the addition of a description of "hate speech" in the proposed new subsection 6(4) of the BOSE that contains a non-exhaustive list of grounds that includes, but is not limited to, hate against a person or group of persons 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". While we support a strong and broad description of hate speech, we would recommend the use of an exhaustive rather than non-exhaustive list of grounds to ensure that the definition remains clear and well-understood. For example, this approach would be consistent with the approach taken by the Racial Discrimination Act 1975 which at section 9 outlaws discrimination against a person on the basis of a specific, exhaustive list of grounds (race, colour, descent or national or ethnic origin).

Snap recommendation 5: We recommend that the list of grounds of "hate speech" in the proposed new subsection 6(4) of the BOSE be an exhaustive list rather than a non-exhaustive list.

Transparency requirement

We support the inclusion at clause 15 of the Amendment Determination of a new additional expectation that service providers publish at least annual transparency reports providing a range of information that is specific to Australia (unless it is not reasonably practicable to do so). Snap currently publishes transparency reports twice a year to provide insight into Snap's safety efforts and the nature and volume of content reported on our platform. This includes proactively releasing a standalone page with data specific to Australia.

However, while the Amendment Determination provides some general high level requirements around what transparency reports should cover, we recommend that it also clarify the Government's expectations around the detail required to be provided, noting that currently all online service providers would be subject to the expectation. For example, a possible 'black letter' interpretation of the proposed new paragraph 18A(1)(a) of the BOSE would require each provider to provide granular data around a service's enforcement of each relevant terms of use, policy, procedure and standards of conduct. We therefore recommend that the proposed new transparency requirement (or accompanying regulatory guidance) specify that the expectation should be flexibly applied and that providers are not expected to publish highly detailed or granular reports, although as a minimum they should include key information or data on the amount of harmful content reported and actioned across the major harm categories.

We are also concerned with the specific requirement for providers to include the number of active end-users in Australia (including children) each month during the relevant reporting period. While we understand and support the general requirement for transparency around the number of end-users of a platform in Australia, we note that this is already achieved through the proposed new subsection 20(5) that provides an additional expectation that the eSafety Commissioner may request a report on the number of active end-users of the service in Australia (including children) during a specified period.

Regular reporting on the number of active end-users will be a significant burden on platform providers, while information around the number of end-users in a country will often be commercially sensitive information that providers may not wish to report on. Requiring this information to be included in a transparency report would require that information to be made public (in contrast to the proposed new subsection 20(5) that would only require such information to be provided directly to the Commissioner) meaning that compliant providers would be at a commercial disadvantage against non-compliant providers. Alternatively, we recommend that information on the number of active end-users be able to be provided confidentially to the eSafety Commissioner instead of in a public transparency report.

Notwithstanding the above, should the Government nevertheless decide to keep a requirement for providers to report on their number of active end-users in Australia, whether in a transparency report or confidentiality to the eSafety Commissioner, we believe that this can be achieved through a general requirement rather than a strict requirement for granular monthly data, which can be of even greater commercial sensitivity. For example, consecutive monthly data may be extrapolated into indications, however correct or incorrect, that a platform is slowly gaining or losing commercial market share in the region. A general requirement would instead give providers flexibility in how they report on their number of end-users (eg. by enabling them to report on the number of end-users at the end of or as an average during the reporting period) while still ensuring there is full transparency of the number of a provider's end-users in Australia.

In addition, we also suggest that the proposed new expectation for transparency reports specify that, notwithstanding the range of information included at proposed new subsection 18A(1) of the BOSE, a provider is not required to include in a transparency report any information that the provider already makes public, such as on its website. For example, rather than providing in each transparency report information on the safety tools and processes deployed by the service as required by the proposed new paragraph 18A(1)(b) of the BOSE, which may be detailed, lengthy and repetitive, a provider may wish to instead provide that information more prominently elsewhere on its website, such as through a featured part of its safety centre.

Snap recommendation 6: We recommend that the proposed transparency requirement, or accompanying regulatory guidance, specify that the expectation should be flexibly applied and that providers are not expected to publish highly detailed or granular reports, although as a

minimum they should include key information or data on the amount of harmful content reported and actioned across the major harm categories.

Snap recommendation 7: We recommend that the proposed transparency requirement not include an expectation for providers to report on the number of active end-users of the service in Australia (ie. we ask that proposed new paragraph 18A(1)(d) be removed). Alternatively, we recommend that this information be able to be provided confidentially to the eSafety Commissioner.

Snap recommendation 8: Notwithstanding our recommendation 7 above, should a requirement of any kind to report on the number of active end-users be kept, we further recommend that the proposed new paragraph 18A(1)(d) of the BOSE be changed from "the number of active end-users of the service in Australia (including children) each month during the relevant reporting period" to "the number of active end-users of the service in Australia (including children) during the relevant reporting period".

Snap recommendation 9: We recommend that the proposed new expectation for transparency reports specify that transparency reports are not required to contain any information that the provider has already released and is readily accessible.

Conclusion

We again thank the Department for the opportunity to provide a response to the proposed amendments to the BOSE. The proposed changes will help to ensure that the BOSE keeps up to date with the evolving online environment and we hope that our feedback and recommendations in relation to some of the expectations are helpful for finalising the Amendment Determination.