Asia Internet Coalition (AIC) Comments on

Public Consultation on Enhancing Online Safety For Users in Singapore
Submission to Ministry of Communication and Information

---

10 August 2022

Ms. Josephine Teo,
Minister of Communications and Information,
Ministry of Communications and Information,
140 Hill Street #01-01A,
Old Hill Street Police Station,
Singapore, 179369

Dear Minister Josephine Teo,

Subject: AIC Comments and Recommendations on the Public Consultation on Enhancing Online Safety For Users in Singapore

On behalf of the Asia Internet Coalition (AIC) and its members, I am writing to express our sincere gratitude to Singapore's Ministry of Communications and Information (MCI) for allowing the AIC to submit detailed comments on the Public Consultation of Enhancing Online Safety For Users in Singapore. As an introduction, AIC is an industry association of leading Internet and technology companies in the Asia Pacific region with an objective to promote the understanding and resolution of Internet and ICT policy issues. Our current members are Airbnb, Amazon, Apple, Cloudflare, Expedia Group, Meta, Google, Grab, LinkedIn, LINE, Rakuten, Spotify, Twitter and Yahoo, and Booking.com.

First and foremost, we commend the Government of Singapore's public consultation aimed at enhancing online safety for users in Singapore. On this note, we share the Singapore Government's concerns of the need to put in place measures to protect young users from harmful online content such as child abuse, terrorism, sexual harm, self-harm and public health that can lead to serious consequences in the real world, and are supportive of the overall intention of Singapore's proposed online safety codes. We appreciate MCI limiting the scope of the codes to only apply to designated social media services, particularly those that have high reach and impact, and that the short list of designated social media services will be made publicly available, with the possibility of adding more social media services to the list if they are deemed as high risk at a later stage.

We also appreciate MCI's understanding and openness to work with social media services to take reasonable steps to remove harmful content as soon as reasonably practical depending on the level or type of egregious content identified by IMDA. Allowing reasonable timeframes for takedown that is proportionate to content type enables safety teams to appropriately

prioritize incoming content issues, in particular, prioritizing action on Child Sexual Abuse Material (CSAM) and Terrorist and Violent Extremist Content (TVEC) type content over a less critical content.

As responsible stakeholders in this policy formulation process, we look forward to the transparent and further rounds of consultation in September 2022 once MCI releases its draft codes. We wish to emphasize that the industry stands ready to collaborate and work closely with the Government to adopt an outcomes-focused approach to the drafting of Singapore's online safety codes that allows for innovation and future-proofing. We firmly believe that Codes drafted based on achieving desired outcomes, without prescriptively dictating the means of achieving that outcome, will allow services to design the most appropriate/effective solution for their particular services. Not tying compliance to a specific method/technology will also avoid stifling innovation in technological solutions to safety challenges.

Given the above, please find appended to this letter (in Section A and Section B), detailed comments and recommendations which we would like to respectfully request the Ministry of Communications and Information to consider when formulating the proposed codes of practice to enhance online safety.

Should you have any questions or need clarification on any of the recommendations, please do not hesitate to contact our Secretariat Mr. Sarthak Luthra at Secretariat@aicasia.org or at +65 8739 1490. Furthermore, we would also be happy to offer our inputs and insights on industry best practices, directly through discussions and help shape the dialogue for the advancement of online safety in Singapore.

Thank you

Sincerely,

Jeff Paine
Managing Director
Asia Internet Coalition (AIC)

Detailed comments and recommendations

Section A: General Comments

1. Definition of social media services

- **We seek clarity on how social media services will be defined, and how "designated social media services" will be specified.**
- We propose the following definition: "Social media service" means an internet-based service or application that has users in [Singapore] and that meets all of the following criteria:
  1. The primary purpose of the service or application is to connect users and allow users to interact with each other within the service or application.
  2. The service or application allows users to do all of the following:
     a. Construct a public or semipublic profile within a bounded system created by the service or application.
     b. Populate a list of other users with whom an individual shares a connection within the system.
     c. View and navigate a list of connections made by other individuals within the system.
     d. Create or post content viewable by other users.
- **We recommend the adoption of a risk-based approach, that is based on objective criteria, when identifying "designated social media services".** In scope services should have the opportunity to conduct an individual risk assessment - based on their reach, exposure and existing mitigations - to determine whether any obligations are required.

2. User Safety

- Social media services already do provide users with tools and options to manage their own exposure to unwanted content and interactions, and **we support the proposal that designated social media services provide safety information that is easily accessible to users, including having relevant safety information be pushed to users that search for high-risk content, such as those related to self-harm and suicide. It is important to recognize that all social media services operate differently, serving different user communities, with different risk profiles. Therefore, MCI/IMDA should avoid prescribing specific tools and mechanisms (such as mandatory default settings) that may not be fit for purpose and lead to unintended consequences.**

3. Content categories and clearly distinguishing between material that is 'illegal' and

material that is ‘legal but harmful’

- We recommend making **a clear distinction between what material is 'illegal' and material that is 'legal but harmful', with different levels of obligations for both categories so as to provide clarity to services and maintain consistency between online and offline.**
    - It is appropriate to require the removal of illegal content (such as the Child Sexual Abuse Material (CSAM) and Terrorist and Violent Extremist Content (TVEC) categories mentioned, and others ideally identified as such by the Government through a notice and takedown scheme). However, material that is 'legal but harmful' (to either adults/children) should be dealt with through services’ individual clear and transparent policies. Schemes that require the removal of content that is easily identified as illegal (such as CSAM and TVEC) can be operationalized by companies (for example by using hash sharing databases). It is also incredibly challenging to adjudicate ambiguous or nuanced content categories in the social media context. Adopting the test of illegality for prohibition also maintains consistency between the online and offline world.
    - Proactive detection and removal by services should also be limited to known CSAM and TVEC, as there are challenges with identifying new material. **We recommend that government and private sector should work together to help identify CSAM, TVEC material to improve the ability of services to proactively detect and remove such material using technology.**

4. Government takedown notices

- Referencing para 24, **we propose a notice and take down approach towards specified harmful content on social media content, which requires services to take reasonable steps to remove social media content as soon as reasonably practical**. **Justification on why the specified social media content is deemed as harmful and in public interest to be taken down, as well as details of the content’s online location, should also be provided as part of the government’s take down notice to social media services.**
    - Reasonable timeframes for takedown that are proportionate to content type allows safety teams to appropriately prioritize incoming content issues, and enable them to prioritize action on TVEC/CSAM type content over a less critical content.
- **The abovementioned notice and take down approach should also be applied to specified online accounts that IMDA deems to be harmful. For online accounts, we recommend that MCI/IMDA have a higher threshold given the larger impact on speech, such as the account demonstrating**

recidivist behavior, or posting high severity content. Similarly, justification on why the specified online accounts have demonstrated a high threshold of harm and therefore in public interest to be taken down, as well as details of the specified online account, should also be provided as part of the government's take down notice to social media services.

5. Default child settings

- Social media services that allows users below 18 already have youth-focused safety settings as an additional safeguard for young users. We recommend that MCI/IMDA avoid prescribing tools and mandatory defaults to enable social media services to develop tools appropriate to their risk profiles and evolve those tools over time.

6. User Reporting and Resolution, and Accountability

- We support the proposal for designated social media services to provide an efficient user reporting and resolution process that enables users to alert these services to content of concern. We are also supportive of the proposal that services should, as part of the process, assess and take appropriate action on user reports in a timely and diligent manner.
- **We recommend that transparency reporting obligations should be simple, easy to adopt and scale, and meaningful .** Transparency reporting obligations should be easy for social media services to adopt and scale at reasonable costs. There should also be scope to provide a supporting narrative to any requested metrics, as context is essential in order to derive meaning from content moderation statistics. Additionally, transparency reporting should be outcomes-focused and answer a critical and relevant question that the regulation aims to solve, rather than on specific mechanisms that may only be a small piece of the overall content moderation process.

Section B: Responses to MCI's Email Feedback Form

1. **We propose for social media services to put in place system-wide processes to reduce users' exposure to harmful online content for specified categories of harmful content such as (i) sexual content; (ii) violent content; (iii) self-harm content; (iv) cyberbullying content; (v) content endangering public health and (vi) content facilitating vice and organised crime (para 10 and Annex A). Do these categories cover the**

range of harmful online content you are concerned about? What other areas of harmful online content would you propose to include?

- We share the Singapore government's concerns of the need to put in place measures to protect young users from harmful online content such as child abuse, terrorism, sexual harm, self-harm and public health that can lead to serious consequences in the real world, and are supportive of the overall intention of the proposed online safety codes.

- However, we recommend making **a clear distinction between what material is 'illegal' and material that is 'legal but harmful', with different levels of obligations for both categories so as to provide clarity to services and maintain consistency between online and offline.**

    - It is appropriate to require the removal of illegal content (such as the Child Sexual Abuse Material (CSAM) and Terrorist and Violent Extremist Content (TVEC) categories mentioned, and others ideally identified as such by the Government through a notice and takedown scheme). However, material that is 'legal but harmful' (to either adults/children) should be dealt with through services' individual clear and transparent policies. Schemes that require the removal of content that is easily identified as illegal (such as CSAM and TVEC) can be operationalized by companies (for example by using hash sharing databases). It is also incredibly challenging to adjudicate ambiguous or nuanced content categories in the social media context. Adopting the test of illegality for prohibition also maintains consistency between the online and offline world.
    - Proactive detection and removal by services should also be limited to known CSAM and TVEC, as there are challenges with identifying new material. **We recommend that government and private sector should work together to help identify CSAM, TVEC material to improve the ability of services to proactively detect and remove such material using technology.**

2. **Are you aware of existing safety measures and tools provided by social media services? If yes, which of these measures and tools have you found useful? If no, how do you think social media services can raise awareness of their safety measures and tools?**

- Social media services already do provide users with tools and options to manage their own exposure to unwanted content and interactions, which may include **having relevant safety information be pushed to users that search for high-risk content, such as those related to self-harm and suicide.**

3. **What are your views on the safety measures and tools (paras 13 to 14) proposed to reduce exposure to harmful online content for Singapore-based users? Are there other measures that should be included?**

    - Social media services already do provide users with tools and options to manage their own exposure to unwanted content and interactions, and **this may include having relevant safety information be pushed to users that search for high-risk content, such as those related to self-harm and suicide.**

4. **A key area of concern is protecting young users from the impact of harmful online content. Do you think the additional safeguards proposed for young users (paras 16 to 18) are appropriate and adequate? Are there other safeguards for young users that should be provided?**

    - Referencing para 17, we support the recommendation that designated social media services that allow users below 18 to have youth-focused safety settings as an additional safeguard for young users. **We recommend that MCI/IMDA avoid prescribing tools and mandatory defaults to enable social media services to develop tools appropriate to their risk profiles and evolve those tools over time.**

5. **What do you think an effective and transparent user reporting mechanism (paras 19 to 21) should require? What do you expect the social media services to do after you have reported harmful content to such services?**

    - We support the proposal for designated social media services to provide an efficient user reporting and resolution process that enables users to alert these services to content of concern. We are also supportive of the proposal that services should, as part of the process, assess and take appropriate action on user reports in a timely and diligent manner.

    - **We recommend that transparency reporting obligations should be simple, easy to adopt and scale, and meaningful.** Transparency reporting obligations should be easy for designated social media services to adopt and scale at reasonable costs. There should also be scope to provide a supporting narrative to any requested metrics, as context is essential in order to derive meaning from content moderation statistics. Additionally, transparency reporting should be outcomes-focused and answer a critical and relevant question that the regulation aims to solve, rather than on specific mechanisms that may only be a small piece of the overall content moderation process.

6. **What can the community, the private sector and the Government do to enhance online safety for Singapore-based social media users? What are potential areas of collaboration?**

- **There is room for collaboration between the private sector and government on an outcomes-focused approach to the drafting of the codes that allows for innovation and future-proofing.** Codes drafted based on achieving desired outcomes, without prescriptively dictating the means of achieving that outcome, will allow services to design the most appropriate/effective solution for their particular services. Not tying compliance to a specific method/technology will also avoid stifling innovation in technological solutions to safety challenges.
- We also recommend that the government and private sector work together to help identify CSAM and TVEC material to improve the ability of services to proactively detect and remove such material using technology.