Advice to the Minister for Communications

19 June 2025

Table of contents

Background

This advice is provided in response to a request by the Minister for Communications under section 63C(7) of the *Online Safety Act 2021* (**the Act**).

In providing this advice, eSafety has drawn from a broad evidence base, which I would be pleased to provide in more detail. I have considered the object of the social media minimum age (**SMMA**) obligation as stated in section 63B of the Act and the overarching policy intent of legislative rules (**the Rules**) as set out in the Explanatory Memorandum.

It is my understanding the overarching intention of the SMMA obligation is to protect Australian children under 16 from the risk of harms associated with social media platforms, with a particular focus on content, features and experiences that are detrimental to their safety, health and wellbeing. I understand the intention of the Rules is to narrow the definition of 'age-restricted social media platform' to target the services causing the most harm to age-restricted users, while ensuring children under 16 retain access to services which predominantly provide beneficial experiences.

My advice identifies five possible options which may assist in further aligning the draft Rules with this intention. The advice is structured so that options 1 and 2 address the questions in your request and options 3, 4 and 5 aim to provide longer term options for your consideration. I believe these options would make the draft Rules more capable of promoting the safety, wellbeing and digital rights of children through greater clarity and fewer compliance and enforcement challenges.

It is critical the Rules are made as soon as possible to ensure clarity for industry and the public about which services will need to comply. Delays may result in over-capture of services, potentially reducing children's access to important and beneficial online services. In having regard to this advice, I recommend you prioritise your consideration of options 1 and 2, noting I have provided alternatives to options 3 and 4, and option 5 is prospective.

The purpose of the Act and the draft Rules

Section 63B of the Act states the object of the SMMA obligation is to reduce the risk of harm to children under 16 from certain kinds of social media platforms. eSafety understands the intention is to mitigate:

- The risk of exposure to harmful content, including content that is detrimental to mental and physical health such as suicide, self-harm, disordered eating and sleeping, and substance use.
- The risk of exposure to experiences that are harmful or detrimental to health, including experiences beyond a child's neurocognitive development and maturity.

- The risk that social media can lead to excessive screen-time, social isolation, low community engagement, sleep interference, poorer educational outcomes, poor mental and physical health, and low life-satisfaction.

eSafety understands the Rules seek to provide an exclusion for services that have a lower risk of these harms, and offer benefits such as supporting connection, learning and health.

There are a range of other harms which children may encounter online. These include cyberbullying and various forms of sexual exploitation and abuse, including grooming and sexual extortion.

While the SMMA obligation may reduce these harms on the platforms that are captured, eSafety understands this is not the primary focus. Instead, these harms will continue to be addressed primarily through eSafety's existing complementary regulatory schemes (including our cyberbullying and image-based abuse reporting schemes), as well as relevant Industry Codes and Standards. Potential reforms following on from Ms Delia Rickard PSM's Statutory Review of the Online Safety Act 2021 will also provide an opportunity to consider whether any of these existing schemes should be strengthened. For example, if the SMMA obligation results in cyberbullying and image-based abuse migrating to messaging services that are carved out under the Rules, eSafety will need additional regulatory tools beyond content removal to assist victims and remediate harm.

As a result, while this advice mentions these harms, it does not include a thorough assessment as to the risk of these harms on the services the draft Rules seek to exclude.

The options in our advice – particularly option 2 – seek to confirm and clarify the risks and harms that the SMMA obligation aims to address to promote a shared understanding across government, industry and the public.

Protecting children from online harms on social media

There is mounting evidence to suggest certain design choices, features, and functionality may contribute to or amplify the risk of unwanted and excessive use, and the risk of encountering harmful content or experiences (including enabling highly idealised and edited content as well as other forms of high-risk content or activity). To protect children from the risk of these harms, the Rules should account for these choices, features and functionality.

Currently, the Rules seek to do this by reference to a service's purpose, likely based on the premise that services with listed purposes (such as messaging or gaming) are less likely to have some of the features and functionality which have been associated with harm on social media.

However, based on eSafety's review of online services, some services that may be carved out by the draft Rules utilise the same design choices, features and functionality associated with relevant harms on 'traditional' social media. For example, some online gaming services have design features and functionality associated with harms to health and problematic use, including but not limited to, engagement prompts (such as in-app, push and visual notifications), gamified engagement features (such as badges, levels, or rewards tied to repeated access and engagement) as well as other design features that may be designed to keep end-users on the platform for as long as possible.

Likewise, some messaging services include features and functionality associated with these harms, such as ephemeral content that is only accessible for a short window of time, quantitative social metrics (such as likes, reactions), engagement prompts (such as notifications, reminders, or gamified incentives), geolocation features, as well as appearance editing functions that may contribute to body image issues.

As services continue to evolve, we may see an even greater convergence in the design choices, features and functionality that are offered across services that claim to serve different purposes. We may also see that the way people use services in practice over time diverges from the intended purpose of those services. Online services that may appear low risk today could be misused or repurposed for nefarious aims, therefore presenting a higher risk in the future.

As a result, if a service is excluded based on its 'sole', 'primary' or 'significant' purpose alone, despite the presence of harm, then the Rules may not achieve their intended outcome of reducing risk to children.

The options I propose in my advice seek to mitigate harms associated with social media design choices, features and functionality. Underpinning this advice is eSafety's commitment to fostering systemic change and promoting Safety by Design, encouraging services to consider risks, mitigate harms and embed user safety into all aspects of service design, development and deployment. The options reinforce that the obligation falls to service providers to actively commit to, and implement, safeguards for young users in all aspects of service design.

Advice on options

The following detailed advice sets out the rationale and evidence base for five possible options to make the draft Rules more capable of promoting the safety, wellbeing and digital rights of children.

Option 1: Remove YouTube from the draft Rules, and avoid naming specific services to future-proof the Rules

Naming specific services (e.g. YouTube) in the Rules risks creating inconsistencies with the SMMA obligation's intention to reduce harm to children. Services frequently change their safety practices as well as their features and functionalities, which can alter their risk profile. Accordingly, an exclusion for a named service, such as YouTube, may be inconsistent with the intention underpinning Part 4A of the Act.

While YouTube has many educational and otherwise beneficial uses, eSafety is concerned that the popular use of YouTube among children coupled with reports of exposure to harmful content and the platform's use of certain features and functionality is not consistent with the purpose of the SMMA obligation to reduce the risk of harm.

Results from eSafety's recent Youth Survey indicated YouTube was the most popular social media platform[1] children had ever used, with 76% of 10 to 15-year-olds having used YouTube, making it significantly more popular than other social media platforms such as TikTok, Instagram, and Snapchat, especially among the 10 to 12-year-old cohort.

Among a subset of children who had ever seen or heard potentially harmful content online, 37% reported their most recent or impactful experience with this content occurred on YouTube. Similarly, among a subset of children who had ever seen online hate, 21% reported their most recent or impactful experience of seeing online hate occurred on YouTube.

In addition, recent findings from the Black Dog Institute showed an association between higher daily hours spent using YouTube and greater symptoms of depression, anxiety, and insomnia.[4]

YouTube currently employs persuasive design features and functionality that may be associated with harms to health, including those which may contribute to unwanted or excessive use (such as infinite scroll, auto-play, qualitative social metrics, and tailored and algorithmically recommended content feeds). Separately and combined, these features may encourage excessive consumption without breaks and amplify exposure to harmful content. These design features and functionality, alongside short-form video content, are also widely used on services like TikTok and Instagram, which I understand are intended to be captured by the SMMA obligations.

[1] 'Social media' was defined in the survey as 'any online platform or app where people can both interact with other people and post or share content like photos or videos'. Platforms considered social media for the purposes of this survey were: YouTube, TikTok, Instagram, Snapchat, Facebook, Pinterest, Steam, Reddit, Twitch, X (Twitter), BeReal, Threads, and 'another social media platform or app'. This definition of social media does not necessarily align with the definition of social media in the Act and should not be relied upon for determining which platforms are or are not included under Part 4A of the Act or the draft Rules.

Given the known risk of harms on YouTube, the similarity of its functionality to other online services, and without sufficient evidence demonstrating that YouTube predominately provides beneficial experiences for children under 16, providing a specific carve out for YouTube appears to be inconsistent with the purpose of the Act.

Moreover, the SMMA obligation is limited to preventing children from having accounts. If YouTube is not excluded, nothing in the Act precludes children from continuing to access YouTube (or any other service) in a 'logged out' state.

While YouTube restricts access to certain content, features and functionality in a logged out state, there are certain safety features for accounts that belong to children that can only be utilised in the logged in state. For example, children can be part of a supervised account where parents set viewing restrictions based on age-appropriateness. Therefore, the safety implications of applying the SMMA obligation to YouTube are likely to be mixed, reinforcing the simultaneous importance of online safety education and awareness raising.

In general, I caution against excluding particular services without conditions in the Rules. A legislative instrument excluding a particular service would be based on a point-in-time assessment of that service. This assessment could quickly become outdated if the service introduces new features, functionality or practices that could affect its safety for children. For example, the *New York Times* reported on 9 June 2025 that YouTube has recently 'loosened' its content moderation policies of videos.[2]

Option 2: Clarify certain matters in the explanatory statement to avoid future enforcement challenges

Including certain matters in the explanatory statement will support a shared understanding of the intention and application of the Rules and avoid potential compliance and enforcement challenges. This includes guidance on:

- The specific harms the SMMA obligation and Rules seek to address.
- How to apply the different purpose tests across the Rules, particularly how much weight to give a service's self-described purpose, and what other evidence may be considered – including design choices, features and functionality related to the relevant harms, and user preferences.
- The intended scope of the exclusion for services that have the sole or primary purpose of enabling end-users to play online games, including whether this exclusion also extends to ancillary services like in-game chat or voice communication.

[2] Grant, N., & Mickle, T. (9 June 2025). YouTube loosens rules guiding the moderation of videos. *The New York Times*, accessed 17 June 2025.

Clarity on relationship between risk of harm and purpose

Access to online environments can provide a range of benefits for children, including opportunities for belonging, self-expression, creativity, learning and entertainment.[3] Online services also provide crucial help-seeking avenues for those experiencing distress. For example, among children in Australia aged 8 to 17 years, 1 in 3 (32%) had sought emotional support online in the past year, with 13% indicating they had done so weekly or more often.[4]

Exclusions for services enabling communication, online gaming, and those that support health and education can benefit children by fostering positive online experiences and allowing them to actively participate in the digital environment. However, as noted above, those services may also carry risks of various types of harm.

Confirming the types of online harm the SMMA obligation seeks to address in the explanatory statement and articulating how excluded services minimise the risk of those harms and provide a predominantly beneficial experience to children will provide clarity for industry and the public. This could include identifying which kinds of online services are intended to be captured by each exclusion, for the avoidance of doubt.

This approach would minimise the potential for age-restricted social media platforms to challenge eSafety's compliance and enforcement efforts on the basis that it has misinterpreted the policy intent of the Rules.

'Sole', 'primary' and 'significant' purpose

The draft Rules rely on terms like 'sole', 'primary,' and 'significant' purpose without defining them. There is little guidance on the application of the relevant statutory tests and interpretation of 'sole or primary purpose' and 'significant purpose' in this context. This creates uncertainty for industry and the public, and enforcement challenges for eSafety if age-restricted social media platforms are able to dispute our interpretation of the purpose tests and claim they fall within an exclusion.

Many online services have multiple purposes, and these purposes may change over time. In addition, the way a particular service classifies or markets itself may or may not reflect community understanding and usage, and may not be consistent across various contexts or forums.

---

[3] National Academies of Sciences, Engineering, and Medicine. (2023). *Social media and adolescent health.* Washington, DC: The National Academies Press. https://doi.org/10.17226/27396

[4] eSafety Commissioner. (2022). *Mind the Gap: Parental awareness of children's exposure to risks online*. Aussie Kids Online. Melbourne: eSafety Commissioner.

For example, the Snapchat app is currently categorised as a 'Photo and Video' app on the Apple App Store and as a 'Communication' app on the Google Play store, and has various features and functionality associated with social media platforms. X (formerly Twitter) was categorised as a 'Social' app on the Google Play Store as recently as March 2025, but is now categorised as a 'News & Magazines' app on the Google Play Store, and as 'News' on the Apple App Store. Without clear guidance on the extent to which a service's own statement as to its 'sole', 'primary' or 'significant' purpose is determinative, services may engage in 'regulatory arbitrage' to avoid the SMMA obligation.

The way a service is used in practice – particularly by children – does not always reflect the service's intended purpose. For example, the Saudi Arabian app Sarahah was originally intended for workplace use to facilitate anonymous feedback between employees and employers. Despite its business-oriented design, the app's anonymous messaging feature was widely adopted by children, exposing them to unmoderated content and cyberbullying. Similarly, a recent article from the *New York Times* highlighted the discrepancy between the intended purpose of Instagram, focused on photo-sharing, and the way 'Gen Z' uses the app, primarily for direct messaging and short-form, ephemeral videos.[5]

eSafety notes that draft Rule 5(1)(d) excludes services that are ***used*** 'solely or primarily for business or for professional development'. Unlike the other classes of excluded services, this definition does not rely on a service's intended purpose but rather how the service is used. Noting the delineation between 'purpose' and 'use' in the Rules, it would be helpful for the explanatory statement to clarify how much weight should be given to a service's intended and actual use – and particularly how a service is used in practice by children – in determining a service's 'sole', 'primary', or 'significant' purpose in the other draft Rules.

In sum, eSafety recommends the explanatory statement provide guidance on the different purpose tests; note that a service's purpose may change over time; discuss how much weight to give a service's self-described purpose; and outline some other evidence eSafety may wish to consider in assessing purpose, including how a service is used in practice and the design choices, features and functionality on that service which are associated with relevant harms.

'Sole or primary purpose of enabling end-users to play online games with other end-users'

It would be particularly beneficial for the explanatory statement to provide guidance about the exclusion relating to online games. It is not currently clear to eSafety whether this exclusion is intended to capture services which do not themselves offer games, but rather, offer ancillary features and functionality for gaming platforms. Examples include:

[5] Holtermann, C. (12 June 2025) 'Instagram Wants Gen Z. What Does Gen Z Want From Instagram?', *New York Times*. accessed 16 June 2025.

- services that host games created by users (in addition to other content)
- services used by gamers to message, voice call or video call during game play
- services used by gamers to livestream their gameplay to other players
- devices and consoles (including consoles that may have social interaction functionality built into the console)
- information sharing forums or channels pages on information sharing forums where users discuss gameplay.

These features can include, but are not limited to, livestreaming, messaging, invitations to play, or leaderboards. In certain circumstances, the online gaming service may require the user to also use the service providing the ancillary features and functionality to participate in an online game. This highlights the complexity in determining when a service has the sole or primary purpose of enabling a user to play online games.

Option 3: Add criteria for safety measures to mitigate features and functionality associated with harm

No service is immune from being weaponised or misused. An online service purporting to have a positive or beneficial primary purpose does not necessarily mean the service is less harmful or less likely to expose children to online harms, particularly where the service is not designed with safety in mind.

For example, eSafety's recent Youth Survey highlighted that many harms observed on social media services are also present and experienced by children using certain messaging and gaming services, though to a lesser extent than social media services.[6] 1 in 3 Australian children reported their most recent or impactful experience of cyberbullying occurred on a communication platform,[7] while 1 in 4 reported recent or impactful cyberbullying while online gaming.[8]

---

[6] 'Social media' was defined in the survey as 'any online platform or app where people can both interact with other people and post or share content like photos or videos'. Platforms considered social media for the purposes of this survey were: YouTube, TikTok, Instagram, Snapchat, Facebook, Pinterest, Steam, Reddit, Twitch, X (Twitter), BeReal, Threads, and 'another social media platform or app'. This definition of social media does not necessarily align with the definition of social media in Part 4A of the Act and should not be relied upon for determining which platforms are or are not included under Part 4A of the Act or the draft Rules.

[7] 'Communication platforms' were defined in the survey as apps or platforms to 'chat with, message, call or video call anyone online'. Platforms considered communication platforms for the purposes of this survey were: Discord; Email; FaceTime; Google Chat; IMO; KakaoTalk; Kik; Line; Messenger Kids; Messenger; Signal; Skype; Telegram; Text messages; Viber; WeChat, WhatsApp; Wickr; 'another app or platform to message, call or chat to people online'.. This definition of 'communication platforms' should not be relied upon for determining which platforms are or are not included under Part 4A of the Act or the draft Rules.

[8] In the survey, online gaming included 'online video games' and 'Voice or text chat in a video game or console'. This definition of online gaming should not be relied upon for determining which platforms are or are not included under Part 4A of the Act or the draft Rules.

The draft Rules also do not currently account for the features and functionality that can cause or contribute to harm. As stated earlier, eSafety has observed that many services, regardless of their purpose, utilise features that are associated with harms to health, such as ephemeral content and persistent notifications and alerts. These also have the potential to be used in harmful ways where they may have a negative impact on children's sleep, wellbeing and attention.

A number of jurisdictions, including the United Kingdom, the European Union and some states in the United States, have adopted an approach focusing on mitigating the risk of certain design choices, features and functionality. This includes identifying certain design choices that are associated with excessive use, encouraging harmful engagement that is detrimental to health, or amplifying or exacerbating content and contact related harms, and requiring services to take steps to address or mitigate these harms.[9]

A potential approach to addressing certain harms in the Rules is to adopt an eventual reform involving a two-pronged test that references features and functionality associated with harm. The two-pronged test could require the online service to meet the existing purpose/use test and also meet a requirement to implement effective safeguards and safety measures if it has any of the features and functionality identified as posing a high risk of relevant harm. The criteria to have safeguards and safety measures for the identified features and functionality would need to be the default setting for all accounts.

Features and functionality associated with harm

Social media and other online services are designed to maximise user reach, engagement duration and time users engage on service, and overall activity on the service. Certain design features or functionality may be intentionally crafted to maximise content consumption by tailoring what users see to align with their interests and attention patterns. These designs often introduce time pressures, foster a sense of urgency and minimise friction to encourage continuous engagement. Additionally, many design choices aim to boost user activity by quantifying popularity, prompting and rewarding interactions, and making it easy to connect, share and participate on the platform.

---

[9] In the United Kingdom, Ofcom has identified a number of features and functionalities as posing a risk of harm for the purposes of providers undertaking a Children's Risk Assessment. In the European Union, Article 34 of the Digital Services Act (DSA) requires providers of 'very large online platforms' to identify, analyse and assess any systemic risks stemming from the design or functioning of their service and systems, including algorithmic systems, and their negative effects on children's' physical and mental well-being (among other issues). Article 28 of the DSA requires providers of all online platforms to put in place measures to ensure a high level of privacy, safety and security for minors (children). The European Commission has released draft guidelines for consultation for Article 28. In California, the Protecting Our Kids from Social Media Addiction Act would make it unlawful for the operator of an 'addictive internet-based service or application', which includes but is not limited to social media platforms, to provide an addictive feed or send user notifications to a child/minor under 18 without parental consent. The New York Stop Addictive Feeds Exploitation (SAFE) For Kids Act has also introduced requirements to deal with certain design choices.

Features that aim to maximise user engagement and activity are commonly referred to as 'persuasive design'.[10] There is concern that, particularly in the context of children, such design prioritises engagement at the expense of user health and safety. Although most design features are not inherently harmful, when they prioritise engagement over safety and wellbeing, are implemented without appropriate safeguards, and lack transparent, rigorous impact assessments, they can contribute to or amplify risks that negatively impact children online.

Determining the unique and specific impacts of individual design features is challenging, as harms may result from the cumulative effect of multiple features, or the way these features are operationalised (such as through embedded reward systems).[11]

Additionally, they can be difficult to examine because of the constantly evolving nature of digital platforms. This complexity is further compounded by the limited availability and transparency of data from online services regarding health impacts. Furthermore, the effects of these design features can vary greatly depending on individual factors, including developmental vulnerabilities and the presence of protective factors within the home environment.[12]

There is increasing concern that the use of persuasive design may cross into the territory of 'manipulative design', exploiting children's under-developed cognitive capacities (such as impulse control or self-regulation) or developmental sensitivities, included heightened responsiveness to social feedback and evaluation. These tactics are likely to have a disproportionate impact on children's health and safety. Particularly concerning are design choices that may undermine a child's autonomy or control of their digital experiences. Common features associated with such risks include:

- personalised and algorithmically recommended content (such as recommender algorithms and content moderation tools)
- endless content feeds (such as auto-play and infinite scroll)
- engagement prompts (such as alerts and notifications)

---

[10] 5Rights Foundation. (2023*). Disrupted childhood: The cost of persuasive design*, 5Rights Foundation, accessed 16 June 2025.

[11] Maheux, A. J., Burnell, K., Maza, M. T., Fox, K. A., Telzer, E. H., & Prinstein, M. J. (2025). Annual Research Review: Adolescent social media use is not a monolith: toward the study of specific social media components and individual differences. *Journal of child psychology and psychiatry, and allied disciplines*, *66*(4), 440–459. https://doi.org/10.1111/jcpp.14085

[12] National Academies of Sciences, Engineering, and Medicine. (2024). *Social Media and Adolescent Health.* Washington, DC: The National Academies Press. https://doi.org/10.17226/27396
Maheux, A. J., Burnell, K., Maza, M. T., Fox, K. A., Telzer, E. H., & Prinstein, M. J. (2025). Annual Research Review: Adolescent social media use is not a monolith: toward the study of specific social media components and individual differences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *66*(4), 440–459. https://doi.org/10.1111/jcpp.14085
American Psychological Association. (2023). *Health advisory on social media use in adolescence*, American Psychological Association, accessed 17 June 2025.

- quantifiable social metrics (such as likes, reacts, follower counts)
- ephemeral and time-sensitive content (such as stories, streaks, engagement rewards, and double ticks)
- emerging AI-driven tools and features (including chatbots and content modifications tools).

The above list is not exhaustive. It does not capture all features that may contribute to harm, nor does it address the full range of design elements associated with risks to children's health and safety. Notably communication features (such as direct messaging, livestreaming, public posting, and group messaging) can also play a significant role in perpetuating or facilitating harm, particularly in the context of unwanted or harmful contact and interactions. This list reflects only a snapshot of currently recognised features and their impacts. Ongoing monitoring and investigation of emerging social media and associated functions remains a critical priority, given that children and young people are often the earliest adopters of new technologies.[13]

Measures to mitigate the risk of certain design choices, features and functionality

To mitigate risks of harm, eSafety strongly encourages the Safety by Design approach. 'Service provider responsibility', 'user empowerment', and 'transparency and accountability' are the key foundational pillars of Safety by Design, meaning the responsibility of safety should never fall solely upon the user. Service providers should examine every feature and design aspect of the service to ensure it minimises risks to children and other users.

The safeguards and mitigation strategies recommended across the literature – particularly in major health advisories and grey literature as cited above – vary in scope and approach. They range from more restrictive measures such as limiting or disabling certain features for children, to design-orientated strategies that prioritise children's safety. These include approaches that support user agency by helping children become more informed, empowered, and in control of their online experiences.

Where features are not entirely restricted, many recommendations call for safeguards that apply broadly across all design elements. Key strategies, many of which could be further developed in the Rules, include principles and practices that ensure all features, functionalities and design choices are aligned with child safety and wellbeing.

This proposed consideration would require a clear and detailed articulation of appropriate safety measures to prevent regulatory arbitrage and support effective enforcement. This

[13] Sala, A., Porcaro, L. and Gomez, E. (2024) Social Media Use and Adolescents` Mental Health and Well-being: an Umbrella Review, *Computers in Human Behavior Reports*, *14*(100404), 1–15. https://doi.org/10.1016/j.chbr.2024.100404

could be done by including the ability for eSafety or you to issue directions, from time to time, specifying the required safety measures with the necessary level of specificity.

Context and challenges with this approach that require further thinking

While eSafety believes this option would have the benefit of more closely aligning the Rules with consideration of risks and harms per the intention of the SMMA obligation, we also recognise challenges which are likely to necessitate an alternative approach in the short term, as set out below.

There are complexities in determining when a design choice, feature, or functionality can be harmful and under what conditions. The potential for harm depends not only on individual features and functionality, but also on their strength, influence, discoverability, how they are used, and cumulative effect. The vulnerability and specific circumstances of the child using the online service is also germane to the impact and risk of harm.

In some cases, the evidence on safeguards and best practice advice for certain features is still emerging and may vary to some extent across different types of services. Equally, the intersecting regulatory frameworks applying to relevant content and/or features are still under development. For example, eSafety is currently assessing the industry-drafted Phase 2 Codes, which include proposed measures for social media and other online services to reduce children's exposure to, and empower all users to control their encounters with, 'class 2 material' such as high impact pornography, violence, and themes such as suicide and serious illness, including self-harm and disordered eating. While the Rules could make reference to compliance with related regulatory schemes, such as Industry Codes and Standards as well as the Basic Online Safety Expectations, this may also create additional complexity.

The effectiveness of the approach is highly dependent on how certain features and functionality are defined and/or categorised. If features or functionality are listed, or defined by narrow categories, services may remove one harmful feature only to substitute it with another that achieves the same harmful outcome (for example, removing autoplay but embedding other features that promote continuous use instead).

In addition, a platform's definition and use of features and functionality can vary. For example, TikTok, YouTube, Facebook and Instagram all have short form videos on vertical feeds, with seemingly endless content. However, YouTube and Facebook will automatically move to the next content, while TikTok and Instagram require users to 'swipe'. If a feature is defined narrowly, a service may seek to rely on a small nuance to distinguish its feature. Combined with the constantly evolving nature of services and emergence of new features, the articulation of features would also need to be sufficiently broad to enable some flexibility but not so broad as it would be difficult to implement.

Finally, this approach would require an in-depth assessment whereby platforms must demonstrate to eSafety that they have effectively minimised relevant risks across a range of specified features and functionality. With the rapid pace of change in the technology sector, coupled with the opacity around how features and functionality are operationalised, keeping assessments up to date and attempting to validate the relevant information may create regulatory burden for both services and eSafety.

In light of these challenges and the time constraints to ensure the Rules are made by mid-year, eSafety considers an appropriate alternative to implementing Option 3's two-prong test would be to adopt a combination of Options 2 and 5. This would involve providing guidance about harmful features and functionality in the explanatory statement to the Rules, and monitoring implementation to identify any emerging challenges which should be addressed through further Rules or Digital Duty of Care reforms.

Option 4: Introduce a new rule for lower-risk, age-appropriate services that do not meet the current criteria

There are a number of services that are designed with the intention of providing safer and age-appropriate experiences and content to all users, including young children. These services often promote themselves as offering safer online environments that help children play, learn, and thrive.

Some services of this type may contain highly controlled social engagement features, such as posting content, likes and comments, without providing other common features of social media platforms like direct messaging, video calling, ephemeral content, or appearance editing tools. Many of these services have more robust safety measures, such as the moderation of content before it is posted, strict limitations on what content can be posted, and the provision of terms of use in a child-friendly format.

These services generally present fewer risks of harm to children, with minimised likelihood of exposure to harmful content, contact, or conduct due to the highly restrictive interactivity between users and/or greater levels of content regulation. This aligns with the intent of section 63B, where the risk of online harm is generally considered to be low.

eSafety anticipates some of these services will be excluded from the SMMA obligation under the draft Rules where they have a purpose of supporting education or enabling end-users to play games. However, there may be services which do not meet any of the proposed purpose tests, but are nonetheless safer and beneficial for children to use.

An unintended outcome would be that services designed to provide safer and age-appropriate experiences and content to all users, including young children, could no longer allow children under 16 to have accounts. Consideration could be given to introducing a new

Rule to exclude lower risk, age-appropriate services which have effectively minimised the risk of harm for children of all ages.

Any new Rule that responds to this concern would need to be drafted in a clear, specific, and enforceable way, and further guidance and information would need to be provided in the explanatory statement to align the exclusion with Safety by Design principles and the best interests of the child. These services would need to have effective safeguards in place to protect the health, wellbeing, and broader rights of children.

Services that could rely on this exclusion should include features such as very limited or fully moderated interactivity between users, and high levels of content restriction or moderation (e.g., pre-moderated or curated content designed for young children). Ideally, such services should not have in-app and push notifications, infinite scroll, and short-form video feeds with auto-playing videos switched on by default.

Alternatively, if drafting such a Rule may prove challenging in light of time constraints, eSafety could exercise discretion so as to focus on high-risk services and give less priority to lower risk services that are age-appropriate for children of all ages.

Option 5: Monitor implementation of the SMMA obligation and the Rules for future reforms

As services change and incorporate new features and functionality, so too will their risk. There is a risk children will migrate to excluded services with harmful features, exposing them to the very harms the SMMA obligation seeks to address. This may also have the unintended consequence of children migrating to services where eSafety's current powers to remediate harms such as cyberbullying are less effective.[14]

While I consider option 3 could help address some of these risks, I also acknowledge the complexity of the proposed approach, and that additional time may be needed to fully consider how it could be implemented, including the scope of features and functionalities it would encompass. Given the timing constraints, I suggest you consider revisiting option 3 in future iterations of the Rules or providing further consideration of harmful design choices through complementary regulatory mechanisms such as the proposed Digital Duty of Care being considered as part of broader reforms of the Act.[15] As noted above, this could include

[14] For example, without a power to require services to action accounts in addition to items of content, eSafety will not be able to effectively remediate cyberbullying occurring on services such as messaging services where the online abuse is occurring in closed groups or chats.

[15] Broader reforms to the Act may also enable consideration of how to protect and empower children on services which likely fall outside the scope of the definition of age-restricted social media platform, such as standalone AI companion and chatbot services, which may pose significant risks of harm.

the ability for eSafety or yourself to issue directions from time to time in relation to safety measures or other criteria under the Rules.

To ensure the Rules remain effective and responsive to emerging risks, a process of continuous evaluation and refinement of the Rules will help maintain alignment with the evolving digital environment and uphold the intent of Part 4A in protecting children under the age of 16 from online harms.