



Written Submission of Reddit, Inc. in response to the Statutory Review of the Online Safety Act 2021

Executive Summary

- It is crucial that any changes to Australia’s online safety regulatory framework are reasonable and proportionate to the risks they seek to mitigate, and that they account for the diversity and range of digital platforms that exist. In our submission, we have outlined how these principles can be applied to the online safety industry codes process, transparency requirements and age assurance requirements.
- To that end, we encourage a simplification of the online safety industry codes process enshrined in the Online Safety Act (OSA), to the extent possible. While we are broadly supportive of transparency measures overall, we would caution against any reporting obligations that may require platforms to reveal confidential or sensitive information and give bad actors a roadmap to evade enforcement, or divert resources away from important activities to ensure adherence to legislative requirements, to the detriment of users.
- We caution against the adoption of any regulation that would create unreasonable friction or risk (especially privacy risk) in the online experience for the broader population. As such, any age assurance requirements should seek to minimise the amount of sensitive personal user data that a platform is required to collect and preserve users’ privacy and security.

1. Introduction

Reddit welcomes the review of the Online Safety Act 2021 (the OSA review) and appreciates the opportunity to comment on the matters outlined in the Terms of Reference. We strongly believe that the regulatory framework governing online safety in a country should enhance users’ experience online, while simultaneously protecting them from viewing unwanted content.

Reddit is a medium-sized company headquartered in San Francisco, with a small Australian office which opened in Sydney in July 2021. More than 90% of Reddit’s users are over the age of 18, and Australia forms our fifth largest user base.

Our mission is to bring community, belonging, and empowerment to everyone in the world. To meet this mission and the needs of our users, we employ a multi-layered, democratic approach to content moderation, in which volunteer community moderators, our users and Reddit employees all participate in the safety and content governance of our website. Given this unique community structure, we wanted to use this opportunity to explain Reddit’s distinctive approach to content moderation, before providing our feedback on specific themes raised by the OSA review.

1. Reddit’s unique approach to content moderation

Reddit provides a forum-like platform for people to create self-governing communities of shared interests, known as “subreddits”. There are over 100,000 subreddits on Reddit, which includes



subreddits for people located in the same geographic area (for example, [r/Australia](https://www.reddit.com/r/australia/)¹ and [r/Canberra](https://www.reddit.com/r/Canberra/)²); those who share a common hobby or interest (such as [r/AusFemaleFashion](https://www.reddit.com/r/AusFemaleFashion/)³ and [r/AusBeer](https://www.reddit.com/r/AusBeer/)⁴) or those who may be seeking support, advice or understanding as they undergo a shared experience (such as [r/BabyBumpsandBeyondAu](https://www.reddit.com/r/BabyBumpsandBeyondAu/)⁵, which is a community for both expectant and new parents based in Australia and New Zealand). Due to the unique nature of each community, each subreddit has its own set of specific rules – in addition to Reddit’s site-wide [Content Policy](#) – to help ensure that users engaging with that subreddit have a positive, predictable experience.

a. Upvotes, downvotes, and karma

In contrast to other sites, content on Reddit is primarily ranked and curated by the votes of the users themselves. Any Reddit user with a registered account can vote on each individual post and comment on the site. The content can be voted both “up” and “down” within the subreddit, effectively playing an important role in mitigating against the spread of low-quality content and bad faith behaviour. Downvoted content will be downranked in discussions and on subreddit home pages, while contributions perceived to be of high quality are upvoted, resulting in an increase in visibility. This voting system allows every user on Reddit to participate in the content moderation process at scale, and effectively limits the spread of potentially harmful content.

b. Subreddit rules and volunteer moderators

Sitting in a layer on top of this community voting system are the various rules of individual subreddits. Subreddit rules are both created and enforced by volunteer community leaders (known as moderators, or “mods”), who tailor them to the specific topic of the community (for example, a subreddit whose topic is dogs may have a rule forbidding content about cats). Moderators are empowered to take action against content and users that may breach subreddit rules. Such actions may include removing a post or comment, or banning individual users from posting within the community. They may do so manually, or by using tools such as Automoderator,⁶ which is a simple automated content moderation tool that Reddit makes available to its moderators to configure as they see fit within their subreddit. In this way, mods are effectively core stewards for their subreddit, shaping and influencing the norms, cultures and behaviours of that community.

We have a dedicated Community team at Reddit that proactively engages with moderators to ensure they are supported and equipped with the necessary resources to help them look after their communities. These include resources, such as the [Mod Education Centre](#),⁷ [Mod Help Centre](#),⁸ and [Moderator Councils](#),⁹ where employees engage moderators directly to hear their feedback on everything from safety tools to the Content Policy. Reddit also has a [Mod Code of Conduct](#),¹⁰ which clarifies Reddit’s expectations of mods and provides information to help mods develop subreddit rules

¹ <https://www.reddit.com/r/australia/>

² <https://www.reddit.com/r/Canberra/>

³ <https://www.reddit.com/r/AusFemaleFashion/>

⁴ <https://www.reddit.com/r/AusBeer/>

⁵ <https://www.reddit.com/r/BabyBumpsandBeyondAu/>

⁶ <https://support.reddithelp.com/hc/en-us/articles/15484574206484-Automoderator>

⁷ <https://modeducation.reddithelp.com/>

⁸ <https://mods.reddithelp.com/hc/en-us>

⁹ <https://mods.reddithelp.com/hc/en-us/articles/4415446939917-Reddit-Mod-Council->

¹⁰ <https://www.redditinc.com/policies/moderator-code-of-conduct>



and norms to create and nurture their communities. Violations of the Code are enforced against through various means, and reported on in Reddit’s biannual [Transparency Report](#).¹¹

c. Reddit’s sitewide policies and safety teams

Overarching these networks of subreddit rules and user voting is Reddit’s [Content Policy](#),¹² a set of high-level, principles-based rules set by the company and enforced across the entire site by Reddit’s safety teams. Volunteer moderators are also expected to uphold these rules in their communities. Our Content Policy forbids unwelcome things such as hateful content, harassment, encouraging violence, sharing personal information or intimate imagery without consent, and other behaviour which we feel has no place on Reddit. We are constantly reviewing and updating our Content Policy to ensure that it continues to be fit-for-purpose. For example, we updated Rule 3 of our Content Policy last year, which prohibits the sharing of non-consensual intimate media, to more specifically address AI-generated sexual media.¹³

Reddit’s internal safety and enforcement teams – Reddit employees referred to by our users as “admins” – enforce the Content Policy by responding to user reports and building tools to proactively identify bad behaviour, including by monitoring data signals for evidence of sophisticated commercial spammers and building hashing tools to identify known instances of harmful and illegal content. Our safety teams are also responsible for administering Reddit’s [appeals process](#),¹⁴ which provides users with an avenue to appeal decisions made by admins regarding Content Policy violations. Appeals are evaluated by these teams, and are either granted (e.g., if we determine that the original decision was mistaken) or denied (e.g., if we determine that the original decision was warranted). For mods who feel that content actions taken by Reddit’s internal safety teams were made in error, they may also use [r/ModSupport](#)¹⁵ modmail to submit appeals. Aggregate results of this appeals process are published in our biannual [Transparency Report](#).¹⁶

In this layered, democratic system, content moderation and governance is distributed amongst Reddit users, moderators, and the company, with each sharing some responsibility for the safety of the site. Unlike platforms premised on the promotion of individual user profiles and virality through an influencer/follower model, the purpose of Reddit is to enable communities to have well-moderated, rules-based conversations on the basis of their shared interests. Accordingly, we have designed a community driven, layered approach to content moderation that inherently scales and localises in step with the size and characteristics of our user base, and is distinct in approach from many other digital platforms that currently operate in Australia.

2. Reddit’s Experience with Online Safety Regulation in Australia

Online safety is of the utmost importance to Reddit and accordingly, we have been carefully monitoring and constructively engaging with Australia’s online safety efforts, as well as those of other jurisdictions we operate in. We are a founding Board Member of the [Digital Trust & Safety](#)

¹¹ <https://www.redditinc.com/policies/2023-h1-transparency-report>

¹² <https://www.redditinc.com/policies/content-policy>

¹³ https://www.reddit.com/r/redditsecurity/comments/14sk5ye/content_policy_updates_clarifying_rule_3/

¹⁴ <https://www.reddit.com/appeals>

¹⁵ <https://www.reddit.com/r/ModSupport>

¹⁶ <https://www.redditinc.com/policies/2023-h1-transparency-report>



[Partnership](#),¹⁷ which aims to establish industry best practices that are flexible to a diverse range of company sizes and business models. We were also an active contributor to the development of Australia's first online safety industry codes and provided a submission to the draft Amendment Determination to the Basic Online Safety Expectations consultation (BOSE amendments consultation). We are grateful for the opportunity to contribute to another worthwhile effort to improve Australia's regulatory framework and hope that our feedback to the OSA review will be of use.

Given that the review seeks to assess the operation of the OSA in Australia and whether the current regulatory settings require any changes to help meet the online safety needs of Australians, we thought it would be useful to provide an overview of Reddit's experience in online safety regulation in Australia. As a first principle, we strongly believe that any online safety measure or framework should be reasonable, proportionate and evidence-based, to ensure that policy objectives are achieved without unnecessarily worsening the experiences of users online or burdening smaller platforms with onerous obligations that have little positive impact on user safety.

Second, it is vital that any regulatory framework accounts for the diversity and range of digital platforms, including social media services, that exist. In this respect, we would strongly urge proportionality to be applied to any regulation such that the risk of harm created by a platform is commensurate with the burden placed by any regulation. For instance, adopting a tiering approach or the application of thresholds to definitions would help ensure that the different business models and services offered by platforms are accounted for and that requirements do not inadvertently prejudice smaller companies and hurt their ability to innovate and compete.

A. Reddit's experience with the online safety industry code

As mentioned above, Reddit was an active contributor to the drafting of the Codes of Practice for the Online Industry relating to Class 1A and Class 1B material (Class 1 online safety codes), a process led by DIGI and Communications Alliance and overseen by the Office of the eSafety Commissioner (eSafety). Although Reddit is not a member of either group, we are grateful for their efforts on an extremely complex process and appreciate the extent of collaboration between the industry associations and eSafety.

Given the importance of the codes to online safety regulation in Australia, we thought that it would be useful to take this opportunity to comment on the overall codes drafting process. As mentioned, the industry associations worked hard to ensure that the voices of all platforms were heard and that the codes drafting process operated as smoothly as possible. However, from the perspective of a small-to-medium sized platform, we found the industry codes process complex and cumbersome.

First, part of the complexity to the codes related to the fact that key definitions were tied to the National Classification Scheme (NCS). While the NCS makes sense within the context of traditional media such as film, publications and video games, it was ill-equipped for the diversity of online services and nuanced nature of platforms that the NCS was meant to apply to. We understand that the NCS is currently being reviewed and would encourage amendments to the obligations surrounding industry codes in the OSA to account for the nuanced nature of platform services.

¹⁷ <https://dtspartnership.org/>



Second, in our experience, both in Australia and overseas, industry-led code-development processes are often complex and require large amounts of resources and time to ensure that they are fit-for-purpose. For smaller companies that do not have the resources to dedicate exclusively to the code-drafting process, this can be particularly burdensome. The code-making process must take into account the nuanced and diverse nature of platforms that exist online, such that a one-size-fits-all approach should not be applied to all platforms. To this end, we would encourage the adoption of a code development process led by the regulator, with adequate opportunity for consultation with a wide variety of stakeholders (including civil society and advocacy groups). This would eliminate the opportunity for regulatory capture that currently exists for large companies to dominate the process by virtue of the resources that they can dedicate to it, and would put civil society on more equal footing in the process as well.

B. Efforts to Increase Transparency

At Reddit, we value transparency in our operations and interactions with our users. This is reflected in our biannual [Transparency Report](#),¹⁸ a publication that includes extensive statistics about actions taken by Reddit's safety teams (including content removed by Reddit employees, user account sanctions imposed by Reddit, subreddits banned, the method of detection of violations [human reports vs. automation], and a disaggregation of actions taken by Reddit employees vs. volunteer moderators) and legal requests received from governments, law enforcement agencies, and third parties to remove content or disclose private user data from around the world. In 2023, we established our [Transparency Centre](#),¹⁹ which provides a one-stop shop for Reddit safety, security, and policy information. We also frequently provide updates to users on actions taken to ensure the safety and security of Reddit, including a quarterly safety and security report on the [r/redditsecurity](#)²⁰ subreddit. By using a subreddit rather than a blog post, we are able to engage in an interactive dialogue with our users about issues they are interested in, including answering live questions from our users directly on the platform, rather than a one-way post.

We are proud of our transparency practices, which are bespoke to our platform and therefore provide meaningful information for our community. We are also very supportive of industry transparency initiatives in general. However, consistent with the principles of reasonableness and proportionality in online safety regulation, it is important that any mandatory transparency requirements maintain enough flexibility to ensure that they are actually applicable and make sense for platforms, rather than a one-size-fits-all approach resulting in a compliance race to the bottom. We would also caution that any mandatory obligations on platforms to increase transparency must guard against unintended consequences that may eventuate. For example, as set out in our submission to the BOSE amendments consultation, it is vital that a platform is able to safeguard its own tooling and processes and protect them from bad actors who may use this information to evade identification. Similarly, it is important that any transparency requirements does not oblige a platform to disclose commercially sensitive information, which, particularly for a smaller company operating in a highly competitive environment, could be hugely damaging.

¹⁸ <https://www.redditinc.com/policies/2023-h1-transparency-report>

¹⁹ <https://www.redditinc.com/transparency>

²⁰ <http://www.reddit.com/r/redditsecurity>



We caution Australia against approaching transparency reporting with a one size fits all mentality, as we have seen recently in other jurisdictions. There is a wide diversity in the range of platforms that exist, both in terms of their services and their business models; accordingly, meaningful disclosures look different for different services. Overly prescriptive reporting requirements or transparency templates run the risk of creating products that are at best non-applicable and at worst nonsensical for certain platforms, depending on their structure. We would encourage that this be kept front in mind, and that any proposed changes to the OSA, including in relation to transparency and reporting requirements, account for the nuances of platforms and allow for flexibility in compliance.

3. International approaches to address online safety

Reddit operates globally and as we continue to serve our users worldwide, we are closely watching regulatory developments throughout the world. As such, we appreciate the consideration of international approaches to online safety regulation in the OSA review's issues paper. To optimise regulatory efficiency and minimise burdens on platforms, we strongly advocate for coherence and alignment across regulatory frameworks. In this respect, it is critical that online safety regulation should be reasonable and proportionate to the harms it seeks to address and must be balanced against users' rights to privacy, free expression and access to information.

Like modern society, the Internet is made up of an incredible range of spaces tailored to the individual characteristics and interests of its users. This extends to social media, which is often conceptualised as a monolith when in reality, there are variations and gradients to how social media platforms are built and used, with users of different ages and backgrounds attracted to a variety of platforms for their own particular reasons. In Reddit's case, we are a platform primarily used by, and targeted to, adults, with more than 90% of our users globally being above the age of 18. The age of our primary audience is reflected in our Apple App store rating of 17+, which allows Reddit to limit its download availability to minors and leverage device-level parental controls. This is also supported by third-party research, with the Pew Research Center reporting that most (around 86%) teenagers between 13 and 17 surveyed had never used Reddit.²¹

Still, we recognise that children can, and do, try to access spaces that may not be suitable for them and accordingly, we have tools in place to ensure that minors are able to use our platform safely. Reddit employs age-gating on mature content, requiring users to affirm they are over 18 before they can view it. To ensure that mature content does not occur on non-gated parts of the site, we have both self-tagging of content and communities by users and moderators as well as human and automated tools using machine learning to classify content on the backend. In addition, content and communities that are tagged as 18+ are not discoverable on our platform unless a user has actively affirmed that they are over 18 in their account settings and opted-in to viewing such content. We also take steps to ensure mature communities do not appear in mixed feeds such as r/all or r/popular (the Reddit "home page"), whether a user has opted into adult content or not.

While minors should be protected from viewing content that is not intended for their consumption, we also believe that the Internet is an open space for all, and that users should not be closed off from potentially helpful resources, such as online support communities. To that end, we would caution against the adoption of any regulation that would place unnecessarily onerous obligations on users at

²¹ Pew Research Centre, [Teens, Social Media and Technology 2023](#), 11 December 2023.



large and create unreasonable friction or risk (especially privacy risk) in the online experience for the broader population.

For example, any age assurance solution that requires individual platforms to collect more information and data on users than it already possesses could potentially threaten the safety of users. Accordingly, we believe that any age assurance requirements should seek to preserve users' privacy and minimise the amount of sensitive personal user data that a platform is required to collect.

To that end, we applaud the nuanced approach adopted by the Office of the eSafety Commissioner in its Roadmap for Age Verification,²² which, rather than rushing to recommend or implement regulations to prevent access to certain content, recognised that there are serious concerns about the effectiveness of age assurance mechanisms and associated risks to privacy and security. We also commend the Australian Government for commencing an age assurance pilot, rather than quickly adopting age assurance technologies that may be immature or unsuited to the needs of minors.

In further considering the role of age assurance mechanisms online, the European Commission's "[Mapping age assurance typologies and requirements](#)"²³ research report is instructive. It emphasises that:

[A]ge assurance is a complex matter, and it is far from straightforward as to how it is to be implemented in given situations. That is why age assurance should not be construed as a silver bullet for online child protection. Instead, it should be considered as one of the many tools to protect and further the experiences of children online.²⁴

The European Commission's report also contains a methodology for assessing the necessity of age insurance and determining the method of age assurance, which we think would be useful in considering the design and outcomes of Australia's age assurance pilot:

(1) Proportionality; (2) Privacy; (3) Security; (4) Accuracy and effectiveness; (5) Functionality and ease of use; (6) Inclusivity and non-discrimination; (7) Furthering participation and access; (8) Transparency and accountability; (9) Notification, challenge, and redressal mechanisms; and (10) Hearing the views of children.²⁵

We would urge a considered, evidence-based approach to assessing the appropriate method of age assurance, its impact on websites and users (including minors) and any risks to user privacy. In fact, we consider it important to be nuanced and evidence-based in all efforts to protect adolescents, lest a measure unintentionally hurt some individuals while meant to protect others. A growing body of academic work points to the conclusion that the impact of social media on young people is heterogeneous – that is, it impacts different individuals in different ways. For example, Dutch researcher Patti Valkenburg of the University of Amsterdam notes that while social media's impacts

²² Office of the eSafety Commissioner, [Roadmap for age verification](#), March 2023.

²³ European Commission, [Research report: Mapping age assurance typologies and requirements](#), February 2024.

²⁴ European Commission, [Research report: Mapping age assurance typologies and requirements](#), February 2024, p 9.

²⁵ European Commission, [Research report: Mapping age assurance typologies and requirements](#), February 2024, p 9.



on wellbeing may be deleterious for a small amount (10-15%) of adolescents, it can actually be helpful to the same proportion of others (though for most, the impact is negligible either way).²⁶

Conclusion

We applaud the OSA review's intentions to further refine the OSA and improve online safety regulation in Australia. However, it is critical that any changes to the OSA are reasonable and proportionate to the harm it seeks to address. In addition, any efforts to change regulatory settings should account for the diversity and range of digital platform services that exist online and how they will be affected by these regulations. Broad-brush legislative changes without due regard to the potential impact on users could harm the online experience of many Australians, without necessarily improving their safety outcomes.

²⁶ Patti M. Valkenburg, Adrian Meier, and Ine Beyens, "Social Media Use and Its Impact on Adolescent Mental Health: An Umbrella Review of the Evidence, *Current Opinion in Psychology*, Vol. 44, 2022, p. 58-68; available from <https://www.sciencedirect.com/science/article/pii/S2352250X21001500?via%3Dihub>.