X Corp.

1355 Market St #900
San Francisco, CA 94103

1 July 2024

Strategy and Research
Online Safety, Media and Platforms Division
Department of Infrastructure, Transport, Regional Development, Communications and the Arts
GPO Box 594, Canberra, ACT 2601

For attn: Delia Rickard, PSM

By email: OSAReview@communications.gov.au

Dear Ms Rickard,

Thank you for the opportunity for X to provide a written submission as part of the statutory review (the "Review") of the Online Safety Act 2021.

X Corp. remains dedicated to protecting our users, resisting censorship, promoting transparency and consistently refining our safety measures, policies and tools to better serve our community and we work tirelessly to create a secure environment for our users. Safety and freedom of speech can and must coexist. Our dedicated Safety Team works every day to make X a better, safer place for everyone – users, partners and clients alike - for people across Australia and around the world.

We also remain committed to working with the Australian Government, the eSafety Commissioner (eSafety), our industry partners and wider society as we continue to strengthen our online safety measures and protections, while maintaining our overarching commitment to freedom of expression, privacy, and procedural fairness. We trust this written submission will be a useful input to your review.

Thank you again for the opportunity to provide input to this important process.

Kind regards,

X Corp.

Overview

Since the Online Safety Act 2021 (the "**Act**") first commenced in January 2022, X has maintained that online content regulation requires a proportionate approach to balance protections from harm with human rights and other vital interests, including freedom of expression, privacy, and procedural fairness.

The current implementation of the Act does not properly balance these vital interests. We note that an express question to be considered as part of the Review is whether additional safeguards are needed to ensure the Act upholds fundamental human rights and supporting principles. The clear answer to this question from X Corp.'s perspective is yes, given the way in which operative powers under the Act are exercised by the eSafety Commissioner, with limited consideration as to how to maintain free speech and public debate in discussions around the Act, its implementation and enforcement.

This submission outlines X Corp.'s views on and concerns with the current operation of the Act, and the key areas set out in the Review.

The structure of the Act is duplicative, burdensome, onerous and inefficient

As a primary observation, the Act overlaps with a number of other pieces of regulation, at both Federal and State level, some of which are also undergoing reform. At the federal level, there are at least 17 processes, inquiries and consultations that intersect with the OSA.

A summary of some of the most relevant include the recently announced Government-led and funded Age Assurance testing; the Joint Senate Committee on Social Media and Australian Society; the Government role in the dispute and complaints resolution processes of digital platforms (a recommendation of report 6 of the ongoing ACCC's Digital Platform Enquiry); the Department of Home Affairs report on algorithms on digital platforms; the Government's draft misinformation and disinformation bill; the review of the Privacy Act 1988 expected to be introduced to Parliament in August 2024 including the Government's agreement to implement a Children's Online Privacy Code to promote the design of certain services in the 'best interests of the child' as well as the Government's consideration of dedicated anti-doxxing legislation under the same; the second stage of modernisation of the National Classification Scheme; anticipated draft legislation on hate speech and religious discrimination; Criminal Code Amendment (Deepfake Sexual Material) Bill 2024 introduced; development of Phase 2 Industry Codes of Practice for the Online Industry imminent; commencement of the *Digital Identity Act 2024* (Cth) by 1 December 2024; likely reforms leading to specific obligations on digital platforms to address scams, harmful apps and fake reviews (recommendations of report 5 of the ACCC's ongoing Digital Platform Services Inquiry; several state-based proposals for age restrictions for social media use; reform to electronic surveillance regulation; potential measures to tackle extreme online misogyny; and consultation and consideration of reforms required across a wide range of areas in relation to Artificial Intelligence.

The Act also targets and imposes obligations in relation to a range of content that is already illegal (and very severely penalised, in some cases). It is X Corp.'s contention that serious consideration should be given to whether or not the Act is the appropriate vehicle to regulate the availability of all subject content, or whether pre-existing offences and investigative processes are more appropriate vectors for that regulation, and the relevant government agencies more appropriate parties to carry out that regulation.

By way of example: Child Sexual Exploitation Material (CSEM) is already a reviled and highly illegal form of content. It is also an offense to counsel, promote, encourage or urge the doing of a terrorist act or the

commission of a terrorism offense (which ought to substantially cover 'pro-terror material'). It could not seriously be said that X (or certain other platforms) has (or have) been 'tolerant' of such illegal material in the past. Nevertheless, both CSEM and pro-terror material are brought within the ambit of the Act by virtue of their coverage in the *classification regime*, with those terms *not actually ever being used within the Act itself*. Whilst it may be appropriate to have some regulations mandating how services are to interface with law enforcement when it comes to illegal content, or otherwise regulating processes for dealing with it, imposing further prohibitions on hosting or disseminating that content in the way the Act does is at minimum arguably redundant. This also reflects the critical need for ensuring there is not overlap across the relevant regulations in Australia.

The Act regulates the availability of that illegal content on services under a regime that applies both to illegal and non-illegal content. While some aspects of codes are aimed specifically at CSEM and pro-terror content, most are not, and the regime is vague (on its face, if not in the mind of the regulator) as to the various 'systems, processes and technologies' etc to be employed to address the availability of that content.

This structure means:

(i) the way in which illegal content is dealt with is not subject to an appropriately targeted, specific and harm-based set of practices, informed by the requirements of the enforcement agencies actually charged with dealing with that content; and

(ii) legal, but regulated, content falls to be dealt with under a regime that does not adequately distinguish between it and illegal content, potentially leading that content to be overregulated and/or regulated by stealth in a way out of keeping with the reasonable expectations of the Australian community.

Within itself, the Act then applies multiple layers of regulation to the same conduct/services. This is particularly a concern in the context of the further extension of the BOSE, with the interaction between the BOSE, the Consolidated Industry Codes of Practice for the Online Industry (including, particularly, the Social Media Services Online Industry Code (SMS Code) and the recently registered Online Safety (Relevant Electronic Services – Class 1A and 1B Material) Industry Standard 2024 and Online Safety (Designated Internet Services – Class 1A and 1B Material) Industry Standard 2024, being unclear and potentially problematic. The full implications of these latter instruments are yet to be properly understood given their complexity.

We would refer you to X Corp.'s submission from February 2024 on the draft Online Safety (Basic Online Safety Expectations) Amendment Determination 2023 (**Amendment Determination**) under the Online Safety Act 2021 for further examples of the problematic interactions between these various instruments.

Undefined and/or subjectively defined terminology

Exacerbating the issue outlined above, is eSafety's expansive approach to: (i) the nature of the content in respect of which it considers it has powers under the Act, including material which, we anticipate, ordinary Australians would not otherwise consider to be subject to regulation, such as violent content with obvious news value; and (ii) the exercise of its powers under all regimes, from cyber abuse to BOSE, all of which currently operate in relation to a far wider scope of material than would be anticipated based on the wording of the Act alone.

There is no definition of ‘online harm’ under the Act, meaning that, at a fundamental level, what it is that is being regulated under this regime is not clear. Consequently, it cannot be clear how eSafety is interpreting or classifying material as falling within the ambit of the Act. This is particularly problematic when it comes to eSafety taking appropriate enforcement action and, as a corollary to this, the industry implementing effective proactive measures. ‘Online harm' should be explicitly defined in law, with accompanying *extensive* guidance provided by the regulator to aid industry in interpreting the relevant definitions. Without this, the entire regulatory regime is undermined.

Conversely, and even in the absence of this threshold definition, certain other terms are expressly defined in the legislation.

By way of example, **‘Cyberbullying material’** is defined under the Act to refer to material that:

> *‘would be likely to have the effect of . . . seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child’*

‘Cyber-abuse material’ is defined to mean:

> *‘material that an ordinary reasonable person would conclude was likely intended to cause serious harm to an Australian adult; and an ordinary reasonable person in the position of the targeted Australian adult would regard as being, in all the circumstances, menacing, harassing or offensive’.*

The problem with each of those definitions, however, is that they are each inherently subjective and, as a result, open to differing interpretations. This subjectivity, when coupled with the manner in which eSafety interprets its powers under the Act, and the fact there is no underlying definition of online harm, provides scope for an unreliable legislative regime.

As a final illustration of the same issue, X was troubled by the introduction into the BOSE via the Amendment Determination of the term "hate speech", a term not currently used in Australian legislation. Australian legislation relevantly addresses matters in the nature of hate speech using the terms "racial hatred" or "vilification", which have a much more specific legislative meaning (usually related to racial or religious vilification) than the broad and non-exhaustive definition proposed in the draft Amendment Determination. That definition was of course then removed from the final Amendment Determination but included in the Explanatory statement "*so that guidance could be given outside of the legislative context*". That approach serves only to compound the ambiguity of the term and the uncertainty of its application.

X contends that it is inappropriate for such an ambiguous term to be introduced into Australian law, by way of introduction into the BOSE specifically, thereby having a significant impact on Australia's online landscape, without being first subject to Parliamentary scrutiny and debate.

Without terminology/concepts being embedded in the Act's language, with such terms thereby having been subject to parliamentary scrutiny, and, correspondingly, going on to be interpreted by practitioners and courts in the ordinary way, rather they are arising from processes such as ministerial determinations and legislative instruments (which are themselves somewhat subjective in nature in that they lack the precision of legislation). That then leads into the dynamic which is otherwise being addressed by X in this submission, which is that eSafety promulgates a highly prescriptive interpretation of the obligations. With the instruments being given the force of law under the Act, that interpretation then effectively becomes a

statutorily-prescribed set of technical requirements (unless an industry participant wishes to undertake the expense and risk of litigation arguing the contrary), without having gone through the proper processes.

In the same vein, we recommend the processes surrounding the development of the Phase 2 Industry Codes under the Act (the **Class 2 Codes**) be deferred at least until the Review is completed and its recommendations considered. The Class 2 codes are indicative of the need for such clear definitions and process, given the significant new regulatory and compliance requirements they would add.

Aggressive nature of investigation and enforcement

In the explanatory memorandum to the Act (**EM**) it states:

> *Compliance with the BOSE would be voluntary, although there is an expectation that social media services would generally seek to uplift their online safety practices to best adhere to the new regulations and avoid potential impacts on company reputation.*[1]

In practice, however, eSafety has sought to enforce its interpretation of the BOSE as a set of inflexible obligations on industry participants. A pattern has emerged of eSafety asking very specific, detailed, and lengthy questions, with underlying assumptions about what constitutes purported problems, directed to its own interpretation of the BOSE. Service providers are then required to report their sensitive operations in granular detail, with specific measurements, thereby disclosing confidential, business critical, information, before eSafety unilaterally determines which aspects of the responses to those questions are relevant and should then be made public - including in the face of specific requests from providers that certain information be kept confidential. Providers whose measures may not align with the regulator's preferred approaches are then publicly targeted. All of the above takes place inside a framework which provides no clear procedural or legal supervision of the exercise by eSafety of its purported powers.

With specific reference to the directed questions asked of providers, these go far beyond requiring a report of compliance against some or all of the BOSE, as the Act states - and go far beyond what would be reasonably necessary in order to assist eSafety in understanding the extent to which providers complied with the applicable BOSE. Instead, they constitute invasive inquiries into, and require disclosures of, a provider's operations, which facilitates a practice of "naming and shaming" industry participants, without identifying any specific standards, best practices and/or any other technical recommendations, and without affording the named service providers due process or procedural fairness.

The current operation of the Act is excessively one-sided, and sets eSafety against industry, rather than inculcating meaningful collaboration and consultation between eSafety and industry in order to minimise risks online and to promote online safety for Australians.

It would be in the public interest for the Act to clarify what terms of collaboration, consultation and cooperation between industry and eSafety must be specifically to achieve these aims and to ensure such guidance is efficient, effective, and proportional to any specific business and its users.

---

[1] *Online Safety Bill 2021*, Explanatory Memorandum, p 35.

It would also be in the public interest to clarify the operation of Part 15 of the Act as eSafety's approach to the disclosures those provisions supposedly authorise has been deployed in support of the problematic disclosures described above.

Failure to protect sensitive, business critical information

eSafety's ability to unilaterally determine how to deal with confidential, proprietary, private and security related company information related to measures that companies adopt to mitigate against risks of harm, potentially jeopardises the effectiveness of such measures and places industry in an exceptionally precarious position as it works to meet the interests of users, public transparency and its regulatory obligations, without any mechanism under the Act to protect providers from overreach.

Publicly disclosing such information in transparency reports - including in the face of specific requests from providers that certain information be kept confidential - risks seriously undermining the defenses companies put in place to protect users on their platforms. This position is untenable and should be addressed.

There needs to be a framework and a mechanism to ensure that proprietary, commercially sensitive and confidential information is protected - for reasons which include the need to preserve platform security - and correspondingly restricted from publication, so as to give providers confidence that they can participate in eSafety's processes with the assurance that the information that they share will be kept confidential and not be published.

X would recommend that the Government consider minimum safeguards to address these concerns, such as those comparably found in other new digital services regulation internationally.

These and other structural issues will only be exacerbated by the substantive expansions in the BOSE implemented in the recent amendment to the BOSE Determination. Expansion of the expectations to include *'the best interests of the child'* and other generalised references to the impact of '*business decisions*' on safety will (as explained in X's submissions in relation to those proposed changes) significantly increase eSafety's ability to leverage the substantive reach of the scheme using the expansive interpretation of its powers demonstrated to date.

It also creates the converse of what should be the case - which is an appropriate collaborative relationship between industry and eSafety, working together to minimise risks online and promote online safety for all Australians. At present this instead risks creating an overly censorial regime, negatively impacting how companies operate in Australia, with companies erring on the side of over-enforcement in order to be *'seen'* as compliant with their obligations. This not only defeats the purpose of online safety regulation - to ensure freedom of expression while protecting users from the worst type of content - but could also lead to unreasonable and disproportionate application of both platform policies or local laws.

To be an effective regulator, building trust and collaborating effectively with industry is vital to producing optimal outcomes across the board. To aid industry, regulators should seek to deploy a broad set of levers beyond pure enforcement, reconciling the need for flexibility with regulatory certainty, by providing detailed non-binding guidance on how companies should comply, to supplement principles based rules.

No impactful checks on the exercise by eSafety of its purported legislative powers

As is demonstrated by the above, eSafety has interpreted its powers under the Act broadly and has applied them unpredictably and arbitrarily against platforms, sometimes leading the Commissioner to censor speech that is not against the law.

This is exacerbated by the fact that the Act provides no clear or mandatory process through which industry participants may engage with eSafety's issuance of 'transparency reports' or accompanying commentary, with no clear procedural or legal supervision of the exercise by eSafety of its purported powers, particularly in the context of the BOSE, and a lack of substantive and objective assessment of eSafety's reasons for decisions.

Meaningful and genuine transparency on the part of eSafety, which it frequently calls for from industry, is critically important, particularly in relation to decision-making, both to be an effective regulator and to build trust and collaborate appropriately and effectively with industry.

Additional statutory duties are inappropriate

In our experience, there is a tenuous link between the way the existing regulations are enforced and actual harm, which would suggest that additional statutory duties are inappropriate. The amorphous concept of 'harm' against which the Act guards - which, as we note above, remains undefined - necessarily means that compliance and enforcement practices are frequently remote from what a truly risk or harm-based regime might require. A truly targeted and harm-based regulatory approach requires correction of those existing defects, not the layering of additional statutory duties onto an already duplicative, vague and yet aggressively enforced regime.

Likewise when it comes to the prospect of the introduction of statutory duties of care, should they be introduced, it would be critical that such duties be specific, clear, and workable for industry, accompanied by detailed guidance so as to aid industry in their implementation.

Business disruption sanctions

Given the aggressive and targeted manner in which eSafety has interpreted its powers under the Act to date, the provision of more extensive penal powers must be thoughtfully considered, and set in a framework such that sanctions can only be appropriately applied in cases where where enforcement action does not have the intended deterrent effect and/or the risk of harm from a regulated service is such that it is appropriate for the regulator to take other action. In other words, business sanctions should only be applied in the most egregious of circumstances, and as a 'last resort' when the regulator's full range of regulatory levers has been exhausted and has not shown to produce, or be on the way to producing results.

<<<>>>