26 February 2024

Hon Michelle Rowland MP
Minister for Communications
Department of Infrastructure, Regional Development,
Communications and the Arts
PO Box 594
Canberra ACT 2601

CC: The Online Safety Team
Online Safety Branch - Online Safety, Media and Platforms Division

By email: BOSEreforms@communications.gov.au

Dear Minister Rowland,

Thank you for the opportunity to provide a written submission on the draft *Online Safety (Basic Online Safety Expectations) Amendment Determination 2023* (**Amendment Determination**) under the *Online Safety Act 2021* (OSA).

X is 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 useful and are available for further engagement as part of shared efforts to ensure any amendments achieve their intended outcomes, are feasible and meaningfully enhance protections for Australians online. Please don't hesitate to reach out to us with any questions about this submission.

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

Yours sincerely,

X Corp.

Overview

Since the BOSE Determination 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 Basic Online Safety Expectations (**BOSE**) under the BOSE Determination does not properly balance these vital interests. The Amendment Determination threatens to further unbalance this framework by exacerbating its existing issues.

This submission outlines X's concerns with the current operation of the BOSE Determination and related key areas set out in the Amendment Determination. We thank you again for the opportunity to provide this feedback as part of our shared commitments to our Australian users and partners.

Concerns with the operation of the BOSE that are not addressed in the Amendment Determination

In the explanatory memorandum to the OSA (**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, directed to its own interpretation of the BOSE, before unilaterally determining 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, which is a matter of concern in its own right on which we comment further below - and then publicly targeting those services whose measures may not align with the regulator's preferred approaches, all in a framework which provides no clear procedural or legal supervision of the exercise by eSafety of its purported powers.

This approach unfortunately creates the converse of what should be the case: a collaborative relationship between industry and eSafety, working together to minimise risks online and promote online safety. It is also not consistent with the way in which the BOSE was originally presented to the public, contains no impactful checks on the exercise by eSafety of its purported legislative powers, and fundamentally lacks procedural fairness, given the BOSE Determination and Part 4 of the OSA provide no clear, mandatory or independent process through which industry participants may appeal against eSafety's decisions, including the issuance of "transparency reports" or accompanying commentary.

The terms in which the Amendment Determination proposes to expand the BOSE will exacerbate these issues by providing eSafety with yet greater scope through which to enforce its own interpretation of those expectations.

---

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

Out of scope questions

Section 56(2) of the OSA empowers eSafety to issue non-periodic reporting notices to services in respect of BOSE compliance, stating (emphasis added):

> *(2) The Commissioner may, by written notice given to the provider of the service, require the provider to:*
>
> *(a) do whichever of the following is specified in the notice:*
>
> *(i) prepare a report* ***about the extent to which the provider complied with the applicable basic online safety expectations*** *during the period specified in the notice;*
>
> *(ii) prepare a report* ***about the extent to which the provider complied with one or more specified applicable basic online safety expectations*** *during the period specified in the notice; and*
>
> *(b) prepare the report in the manner and form specified in the notice; and*
>
> *(c) give the report to the Commissioner:*
>
> *(i) within the period specified in the notice; or*
>
> *(ii) if the Commissioner allows a longer period—within that longer period.*[2]

Additionally, the EM states:

> *Social media services would also be expected to produce transparency reports when requested by the eSafety Commissioner, although most large companies are already producing such reports with the appropriately trained staff. The estimated cost to businesses of uplifting online safety practices and producing transparency reports is $178,000 per annum.*[3]

Although the clause in the OSA states that eSafety can determine the manner and form of the report, according to the statute, a notice under this section is to require a report "*about the extent to which the provider complies with*" the BOSE. Additionally, the EM explains that such a report is intended to be comparable to the transparency report that large companies are already producing - as X does.

Contrary to this, however, eSafety has interpreted the BOSE in an overbroad manner, and has issued notices under the above provision requiring answers to specific, detailed and lengthy questions, with underlying assumptions about what constitutes purported problems, which have required service providers to report their sensitive operations in granular detail, with specific measurements, thereby disclosing confidential, business critical, information.

These directed questions go far beyond requiring a report of compliance against some or all of the BOSE, as the OSA and EM state - 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 - and have instead constituted invasive inquiries into, and disclosures of, a provider's operations.

For instance, on 21 June 2023, eSafety provided a non-periodic reporting notice asking detailed, sensitive questions about X Corp.'s business operations in relation to the steps X was taking to minimise "online hate", and to enforce its terms of use and hateful conduct policy (**Online Hate BOSE Notice**). Yet, the BOSE Determination currently does not mention "online hate", and the expectations are not

---

[2] OSA s 56(2).
[3] *Online Safety Bill 2021*, Explanatory Memorandum, p 35.

framed in terms that naturally invite such specific questions. X submits that the applicable legislative framework should avoid information being demanded from private entities relying on subjective interpretations of broadly-cast concepts such as 'harm', in a framework providing no clear procedural or legal supervision of that exercise of purported power or increased protections for users.

The style of questioning adopted by eSafety also fails to account for the fact that private companies measure data and processes in their own unique ways, and eSafety demands the aggregation of data in a way that is different to the form in which providers may maintain it. This is a major issue for providers in the operation of the BOSE. eSafety's demands for information need to account for this, since it is manifestly unfair to expect private companies (especially those that also operate outside Australia) to meet specific reporting requirements as granular as those which have been required by eSafety, when this may or may not be supported by their internal systems, may be irrelevant to understanding or solving certain safety issues durably anyway, and especially in the context that eSafety would intend to make public such information received.

Failure to protect sensitive, business critical, information

The BOSE fail to acknowledge the reality that certain information supplied by providers in response to the questions asked by eSafety in its notices is, by its nature, exceptionally sensitive, business critical, information, the public availability of which would carry the material possibility of substantial and adverse consequences, including the potential risk of exposing providers' internal systems, and leads to enormous privacy concerns, both for provider(s) and their users.

Whilst eSafety acknowledges that certain information supplied may not be suitable for publication, providers are required to defend and protect their position in respect of the confidentiality of that information, by providing submissions as to why it should not be published, with the BOSE effectively putting eSafety into a position where it is able to unilaterally decide which information should be released. In X's experience, eSafety has published such information, even in the face of requests for that information to be kept confidential.

Again taking the Online Hate BOSE Notice as an example, the detailed questions in that notice required X to supply a host of sensitive, business critical, information, including, amongst other things, details of specific tools, safety mechanisms and systems implemented on the X platform, as well as breakdowns of its staff (both in numbers and locations). Throughout the process, X reiterated to eSafety on more than one occasion, both in person and in writing, its concerns about the public release of such information, requesting that it be retained by eSafety in confidence. However, ultimately eSafety was able to unilaterally determine which information it would keep confidential and which should be disclosed - and proceeded to publish X staff numbers and geographic locations, information that X had not itself published before.

Ultimately, eSafety is able to unilaterally determine how to deal with confidential, proprietary, private and security related company information, which 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 BOSE to protect providers from overreach. X submits that this position is untenable and should be addressed.

Accordingly, X respectfully submits that 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.

Lack of procedural fairness

X is similarly concerned about a lack of clear and independent recourse in the BOSE Determination to challenge or appeal against the ultimate unilateral decisions made by eSafety - which includes decisions to publish the sensitive and confidential information given by providers in response to notices - or otherwise to engage with the public statements eSafety makes (or proposes to make) including when publishing its "transparency reports". There is also an inherent lack of procedural fairness for providers in the way such "transparency reports" and associated commentary are prepared and issued. eSafety should not be the sole arbiter of what is and what is not released, as is currently the case.

Again citing the Online Hate BOSE Notice as an example, at the conclusion of that process, eSafety published a "key findings" document, summarizing X's responses, which X was *not* given the opportunity to review prior to publication. In addition, eSafety issued a lengthy and full form "transparency report", only *part* of which X was given the opportunity to review. Each of the publications featured specific commentary regarding X, and X was not provided with any warning or opportunity to review or respond to that specific commentary. Furthermore, eSafety did not share with X any specific standards and/or what it considers to be best practices and/or any other technical recommendations following its report. eSafety also issued a uniformly critical press release, the same day as the key findings and full form report, which included additional commentary from the Commissioner on X's measures and service in light of the sensitive information that X provided in response to the notice. X considers this approach to exceed the appropriate use of eSafety's regulatory powers.

Collaborative relationship with Industry

In effect, eSafety employs an overbroad interpretation of its powers under the BOSE scheme to facilitate 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.

This makes the current operation of the BOSE scheme excessively one-sided, which, as a result, seems to set eSafety *against* industry, rather than inculcating meaningful collaboration between eSafety and industry in order to minimise risks online and to promote online safety for Australians. X submits that it would be in the public interest for the BOSE to clarify what terms of collaboration 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.

Interaction between the BOSE and the Online Industry Codes

Following implementation of the proposed Amendment Determination, the intended interaction between the BOSE and the Consolidated Industry Codes of Practice for the Online Industry (including, particularly, the Social Media Services Online Industry Code (**SMS Code**) (**Codes**) is unclear and

problematic. Provisions of the BOSE and the Code overlap or contradict each other in certain of the amendments as drafted.

The Amendment Determination introduces requirements which duplicate and/or go beyond what is stipulated in the Codes, or which are otherwise inconsistent with them. Duplicative or inconsistent regulatory approaches do not lead to better regulatory outcomes and instead are likely to cause confusion and inadvertent non-compliance while imposing burdens that impede innovation in the online industry.

Contrary to the clear statutory intention, the Amendment Determination invites confusion and uneven regulatory outcomes in requiring more onerous compliance under the BOSE in some respects than under the Codes (which themselves impose extensive and detailed regulatory obligations tailored to specific sections of the online industry). X anticipates the Amendment Determination will draw resources away from the frontline protections of the platform, particularly causing trust and safety teams to split their focus. Industry participants may not be able to provide as effective mechanisms, reporting, enforcement or resolution owing to the need to have regard to broad, duplicative and differential compliance measures in carrying out that work. Such a regulatory scheme would not be fit for purpose.

We set out below examples of overlap or inconsistency with the Code:

- Transparency reporting: The proposed section 18A to be added to the BOSE Determination overlaps with, and goes beyond, the reporting required under the Codes.[4] Why the specific information listed in section 18A(1)(a) to (c) should be included in reporting on both the BOSE and the Codes, and the information listed at 18A(d) be reported in relation to the BOSE specifically, is not apparent.

- Readily identifiable reporting mechanisms: The proposed section 15(2) is an expectation to ensure that the service has clear and readily identifiable mechanisms for reports and complaints about breaches of the service's terms of use, policies and procedures, and standards of conduct is directly comparable to the minimum compliance measures relating to complaints and reporting mechanisms under the Codes.[5]

- Enforcement of terms of use: The proposed sections 14(1A) and 14(2) requires industry participants to take reasonable steps (including proactive steps) to detect breaches, and ensure enforcement, of terms of use reflects minimum compliance measures in the Code relating to enforcing terms of use and breached policies and procedures.[6]

- Resolution of complaints: The proposed section 14(3) requires the provider of the service to respond to reports and complaints under the BOSE and provide feedback on the action taken within a reasonable period of time.[7] In contrast, under the Code, a provider of a Tier 1 or Tier 2 social media service must take appropriate steps to promptly respond to Australian end-users that have made reports and inform the reporter in a reasonably timely manner of the outcome of the report or the complaint.[8] Although similar, with significant overlap between the two

[4] Amendment Determination s 15; SMS Code, Minimum Compliance Measures 32, 33.
[5] Amendment Determination s 14; SMS Code, Minimum Compliance Measures 23 - 25.
[6] Amendment Determination s 11, 12; SMS Code, Minimum Compliance Measures 2, 3, 11, 12.
[7] Amendment Determination s 13.
[8] SMS Code, Minimum Compliance Measure 26.

obligations, the BOSE expectation carries a requirement to take action, while the Code only refers to an outcome. Expecting that an action would be taken in response to every complaint or report assumes that every complaint or report is accurate and requires action to be taken. That is inconsistent with the Code's more reasonable position that it is the outcome of the report or complaint that needs to be provided to the reporter.

Amendment Determination exceeds the appropriate scope of the BOSE

X is concerned that certain of the proposed amendments in the Amendment Determination exceed the purpose and intention of the BOSE under the OSA, including concepts and obligations which should have been subject to parliamentary debate and scrutiny. This is a concern where such expectations and indeed their associated examples are treated by eSafety as firm obligations and where non-compliance affects industry participants' reputations, resulting in financial and operational consequences for industry participants.

For example, the Amendment Determination gives as a proposed example of a reasonable step to ensure that end-users are able to use the service in a safe manner, "*assessing whether business decisions will have a significant adverse impact on the ability of end-users to use the service in a safe manner and in such circumstances, appropriately mitigating the impact*".[9] In the context of eSafety's current enforcement practices, such an express example of an expectation could see industry participants coming under regulatory pressure to disclose internal and potentially highly confidential business decisions and associated considerations. That is a significant expansion of the regulation which previously took as its focus the public impact of business practices. Such an expansion is inappropriate, inconsistent with the intention of the statute, and is the kind of incursion on private rights which ought to be the subject of parliamentary scrutiny and proper consideration. This also places online industry participants into situations of conflict, as companies' obligation to report the impact of its business decisions under the expectation may directly contradict a company's obligations to act in the best interests of the company.

X is also troubled by several proposed amendments which exceed the appropriate scope of the BOSE. In particular, the focus in the Amendment Determination on "hate speech", which is currently not a term used in legislation in Australia.

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 provided by the proposed section 6(4) of the Amendment Determination (emphasis added):

> *…hate speech is a communication by an end-user that* ***breaches a service's terms of use*** *and, where applicable, breaches a service's policies and procedures or standards of conduct mentioned in section 14, and* ***can include*** *communication which expresses hate against a person or group of people on the basis of race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, gender identity, disease, immigrant status, asylum seeker or refugee status, or age.*[10]

[9] Amendment Determination s 4.
[10] Amendment Determination s 4.

That definition is obviously broad and interpretive and likely to pick up a far wider scope of speech than that addressed under existing statutes. Its purpose or usefulness is also unclear given it is predicated on the relevant communication being otherwise in breach of a service's terms.

X contends that it is inappropriate for such an ambiguous term to be introduced into Australian law, and in the BOSE specifically, having a significant impact on Australia's online landscape, without being first subject to Parliamentary scrutiny and debate, particularly in the context that the question of whether additional arrangements are warranted to address online harms, expressly including online hate, not explicitly captured under the existing statutory schemes, is scheduled to be considered as part of the statutory review of the OSA[11]. X supports such a review, and, given the same, questions the rationale for introducing such a broad definition as part of the Amendment Determination.

As a leading service for online discourse, the perception of "hate speech" online affects X as much as any other social media service - and other multivarious services where user generated content may be hosted and shared. The regulator's targeting of X in respect of "hate speech" has been evidenced by the Online Hate BOSE Notice. The introduction of a further non-exhaustive and excessively broad definition under the Amendment Determination would likely see such regulator action increase without any positive impact on online safety.

The fact and likely expansion under the Amendment Determination of that action further demonstrates the dangers posed by a lack of procedural fairness and difficulty industry participants face in addressing and solving their concerns with eSafety in relation to the BOSE. Issues of exceeding scope, inappropriate focus on specific industry participants not-fit-for purpose prescriptive solutions, disproportionate regulation and a lack of procedural fairness threatens to undermine the ability of Government, eSafety, and the online industry to work together to bring about effective regulatory outcomes in the best interests of the public.

<<<>>>

[11] https://minister.infrastructure.gov.au/rowland/media-release/ensuring-our-online-safety-laws-keep-australians-safe