Australian Sex Workers Association

Phone – 02 9517 2577
Post – PO Box 854
Newtown NSW 2042
Head Office – 203/1 Erskineville Road
Newtown NSW 2042
Email – info@scarletalliance.org.au
Web – www.scarletalliance.org.au
ABN – 86 612 112 065 | ARBN – 149 618 137

21 December 2023

Executive Manager, Industry Regulation and Legal Services
Office of the eSafety Commissioner
PO Box Q500
Queen Victoria Building NSW 1230

To the eSafety Commissioner

Re: Draft Online Safety Industry Standards 2024 - (Relevant Electronic Services – Class 1A and 1B Material) and (Designated Internet Services – Class 1A and 1B Material)

Thank you for the opportunity to submit to the consultation on the draft 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.

Scarlet Alliance is the Australian Sex Workers Association. Through our objectives, policies and programs, Scarlet Alliance aims to achieve equality, social, legal, political, cultural and economic justice for past and present workers in the sex industry.

Formed in 1989, Scarlet Alliance, Australian Sex Workers Association, is the national peak sex worker organisation. Our membership includes state and territory-based and national sex worker organisations and individual sex workers across unceded Australia. Scarlet Alliance uses a multifaceted approach to strive for equality, justice and the highest level of health for past and present workers in the sex industry. We achieve our goals and objectives by using best practices including peer education, community development, community engagement and advocacy.

Scarlet Alliance is a leader when it comes to advocating for the health, safety and welfare of workers in Australia's sex industry. Through our work and that of our member organisations and projects, we have the highest level of contact with sex workers and access to sex industry workplaces throughout Australia of any agency. Scarlet Alliance represents sex workers on a number of government and non-government committees and advisory mechanisms.

Scarlet Alliance has been engaged as a stakeholder throughout the development and implementation of the *Online Safety Act 2021*,[1] the establishment of the Office of the eSafety Commissioner, the proposed Restricted Access System Declaration[2] and age verification for online pornography,[3] and the Consolidated Industry Codes of Practice for the Online Industry for Class 1A and Class 1B material.[4] In 2023, we also submitted to the Department of Industry, Science and

[1] Scarlet Alliance, Submission No 36 to the Senate Standing Committees on Environment and Communications, *Online Safety Bill Inquiry* (3 March 2021).
[2] Scarlet Alliance, Submission to the eSafety Commissioner on the *Restricted Access System Declaration (Online Safety Act 2021)* (20 September 2021).
[3] Scarlet Alliance, Submission to the eSafety Commissioner on the *Call for Evidence on Age Verification for Online Pornography* (20 September 2021).
[4] Scarlet Alliance, Submission to the eSafety Commissioner on the *Draft Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material)* (23 March 2023).

Resources consultation on safe and responsible AI regulation in Australia,[5] outlining the importance of sex worker interests in the regulation of automated decision-making and generative AI.

Our strong advocacy in these discussions has been focussed towards ensuring that sex workers are considered and included as the government seeks to regulate the tech sector and 'protect' Australians online, and that esafety interventions and regulatory frameworks do not restrict sex workers' access to online information sharing spaces and advertising platforms.

We look forward to participating in the consultations on the development of industry codes and standards relating to Class 1C and Class 2 material, which will have direct and tangible impacts on the safety, privacy and wellbeing of sex workers across unceded Australia.

Mish Pony
Chief Executive Officer

[5] Scarlet Alliance, Submission to the Department of Industry, Science and Resources on *Safe and Responsible AI in Australia* (26 July 2023).

Introduction

While Scarlet Alliance recognises the importance of preventing the creation, distribution, storage and access to Class 1A (child sexual exploitation material (CSEM), pro-terror material, extreme crime and violence material) and Class 1B (crime and violence material, drug-related material) content, we are concerned that the obligations imposed by the Draft Online Safety Industry Standards for Relevant Electronic Services (RES) and Designated Internet Services (DIS) undermine user privacy and security, inaccurately categorise risks, and encourage the use of unproven AI-technologies.

While the Class 1A and Class 1B Codes and Standards intend to prevent unequivocally harmful and prohibited content, regulators must be mindful that these provisions may generate significant consequences for sex workers, health promotion and harm reduction advocates and LGBTQI+ community. This submission is focussed towards aspects of the draft Standards most likely to generate unintended consequences for sex workers in unceded Australia.

Assurances from the eSafety Commissioner that sex workers will not be unduly impacted by esafety measures must be accompanied by meaningful consultation focused towards regulation that is forward-looking, adaptable and clear in scope, and developing understanding sex workers not as generators of online harms, but as members of the community with lived experience of algorithmic discrimination and rights to online safety and digital access and participation.[6] This collaboration and recognition of sex worker expertise and lived-experience will be vital to the development of comprehensive and viable codes and standards relating to Class 1C and Class 2 Material during 2024.

[6] Lisa Visentin, 'Sex industry ‘not my concern’: eSafety Commissioner defends proposed new powers' *The Sydney Morning Herald* (online, 4 March 2021).

Risk assessment

Are the requirements for risk assessment in the draft Standards targeted at the right services and at the right points in a service's development journey? Are the risk factors appropriate?

Scarlet Alliance is concerned with the risk assessment criteria in both the RES and DIS draft Standards. While the automatic classification of many services within both Standards (defined categories and pre-assessed categories in the RES standard, high-impact DIS, classified DIS and general purpose DIS in the DIS standard) provides clarity for industry on compliance requirements, these risk profiles must be carefully decided based on evidence rather than stigma.

We are concerned that pornography websites and apps are pre-assessed as being 'Tier 1 - High Impact DIS' - meaning that they carry the highest risk profile for access and exposure to Class 1A and Class 1B material. While the eSafety Commissioner's Factsheet on the DIS Standard notes that 'pornography sites which enable end-users to post material for other end-users to access'[7] would attract this classification - in fact all websites and apps providing access to X18+ and R (CAT 2) material are automatically classified as Tier 1.

Scarlet Alliance is not aware of ***any evidence*** to demonstrate that pornography sites and apps as a whole present a greater risk of access and exposure to CSEM, pro-terror and crime and violence material than other types of DIS. Pornography DIS already voluntarily use proven measures to limit young peoples' access,[8] and will also have forthcoming obligations relating to Class 2 content as these are developed.

While some pornography sites/apps may be appropriate for Tier 1 classification - this should be based on specific risk-profile rather than catch-all pre-assessment. Additionally, the obligations in relation to Class 2 material **will be applicable to all** pornography DIS, and unwarranted additional obligations may generate confusion and a disproportionate compliance burden, especially for smaller independent producers.

Recommendation 1: That pornography sites and apps are not pre-assessed as 'Tier 1 - High Impact DIS', as there is no evidence that pornography sites as a whole are a higher risk for Class 1A and Class 1B material. Pornography DIS should be subject to the same risk-based self-assessment as other non pre-assessed DIS.

---

[7] eSafety Commissioner, *Draft Online Safety (Designated Internet Services – Class 1A and Class 1B Material) Industry Standard 2024* (Fact Sheet, November 2023) 8.

[8] eSafety Commissioner, *Roadmap for Age Verification and Complementary Measures to Prevent and Mitigate Harms to Children from Online Pornography* (Appendices to Background Report, March 2023) 34-6.

Draft Relevant Electronic Services (RES) Industry Standard

Detecting and removing ‘known’ material, disrupting and deterring 'known' and new material

Is the technical feasibility exception in the obligation to detect and remove known child sexual abuse material and pro-terror material appropriate? How effective will this obligation be with this exception? Are there other examples of systems, processes and technologies that can detect, flag and/or remove known child sexual abuse material and known pro-terror material at scale, which should be highlighted in the Standards or accompanying guidance?

‘Known’ material

While Scarlet Alliance understands the importance of detecting and removing verified Class 1A and Class 1B material, we are concerned that the language used within the draft standards encourages service providers to utilise untested technologies.

While hash-matching technologies such as PhotoDNA are an acceptably accurate in detecting *known* Class 1A and 1B material,[9] the current wording of sections 20 and 21 effectively mandates that RES implement client-side scanning measures unless the ‘technical feasibility’ exception applies.

Despite assurances that the draft Standards do not oblige ‘companies to design systemic vulnerabilities or weaknesses into end-to-end-encrypted services’,[10] there are no clear protections for end-to-end encryption, and providers are concerned that compliance with the draft Standards (or other mandated client-side scanning) represents an insurmountable threat to user security.[11]

Irrespective of the debate over whether client-side scanning effectively breaks end-to-end encryption, these measures amount to a **significant** change in service-user expectations for privacy and security, particularly in RES offering end-to-end encryption systems. At minimum, the Standard should also oblige services to clearly explain to users the nature and extent of information collected and detection measures used.

---

[9] eSafety Commissioner, *Draft Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024* (Fact Sheet, November 2023) 6-7.

[10] Ibid 7.

[11] See: Josh Taylor, 'Proton Mail founder vows to fight Australia’s eSafety regulator in court rather than spy on users', *The Guardian* (online, 15 December 2023),
Also Greg Noone, 'Is Client-Side Scanning the Future of Content Moderation?', *Tech Monitor* (online, 11 August 2022); and
Lily Hay Newman 'Apple's Decision to Kill its CSAM Photo-Scanning Tool Sparks Fresh Controversy', *Wired* (online, 31 August 2023.

Scarlet Alliance is also concerned that the definition of *known child sexual abuse material* in section 6 includes 'non-governmental organisation(s)...generally recognised as expert or authoritative' in identifying CSEM. While the draft Standard specifically mentions the database maintained by the National Center for Missing & Exploited Children (NCMEC), there is no further clarity on which organisations would be recognised, and who makes this decision. As these databases are (and should be) difficult to access and audit, it is appropriate that the eSafety Commissioner provides transparency as to which organisations can provide this expertise as a safeguard against inadvertent incorrect classification of content.

'New' material

Scarlet Alliance believes that the wording of section 22 of the draft Standard does not clearly give effect to the eSafety Commissioner's intention that the Standard 'does not mandate a particular approach or technology',[12] and recognition from the Office of the eSafety Commissioner during the consultation briefing for these standards that AI detection was one of a suite of tools services could employ to 'disrupt' and 'deter' unverified Class 1A and 1B content.

The obligation to 'implement systems, processes and technologies,' and specifying that these may include 'hashing technologies, machine learning and artificial intelligence systems' and 'systems, processes and technologies that are designed to detect key words, behavioural signals and patterns'[13] creates a clear focus towards AI deployment, and encourages services to use unproven technologies to categorise unverified material, despite admissions from the Office of the eSafety Commissioner during the consultation briefing that AI technologies are not yet sufficiently accurate in categorising unverified material.

The discriminatory impacts of inaccurate AI content-moderation have been widely documented,[14] and have direct consequences for sex workers' income, privacy and ability to share health and safety information.[15] Section 22 must be amended to make clear that AI detection is one *of a suite of tools* that RES must utilise to 'disrupt' and 'deter' new/unverified Class 1A and Class 1B material, and provide examples (either within section 22 or the explanatory statement) of other acceptable accountability measures.

Scarlet Alliance makes the following **recommendations** for amendments to the RES Standard:

Recommendation 2A: That the RES Standard include an obligation that services provide users with clear information on the nature and extent of information collected and detection measures used.

---

[12] *Draft Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024 Fact Sheet* (n 8) 15.

[13] *Draft Online Safety (Relevant Electronic Services – Class 1A and Class 1B Material) Industry Standard 2024*, s 22.

[14] This issue has been raised by sex workers, LGBTQI+ community, fat and body-positive activists, sexual health educators and other advocates for over a decade within the public sphere, see for example: Gianluca Mauro and Hilke Schellmann, '"There is no standard": investigation finds AI algorithms objectify women's bodies', *The Guardian* (online, 8 February 2023); Chanté Joseph, 'Instagram's murky "shadow bans" just serve to censor marginalised communities', *The Guardian* (online, 9 November 2019); and Gabriel Nicholas, 'Shadowbanning Is Big Tech's Big Problem', *The Atlantic* (online, 28 April 2022).

[15] Scarlet Alliance (n 5) 5-7.

Recommendation 2B: That the RES Standard clarify (either in the 'example' section following the definition of 'known child sexual abuse material' in section 6, or in the explanatory statement) which non-governmental organisations are deemed to be 'generally recognised as expert or authoritative' in identifying known CSEM, and which body is responsible for making these determinations.

Recommendation 2C: That sections 21-22 of the RES standard include specific recognition that services are not expected to break end-to-end encryption

Recommendation 2D: That section 22 of the RES standard includes specific recognition that AI moderation is one of a suite of tools utilised to 'disrupt' and 'deter' unverified Class 1A and Class 1B content, and include examples (either in the Standard or explanatory statement) of non-AI accountability measures.

Investment requirement for large relevant electronic services

Do you agree with the monthly active user threshold for the investment obligation? Are there other appropriate thresholds that should be considered to ensure the obligation is proportionate to the size and reach of the relevant electronic service?

While Scarlet Alliance believes the active user thresholds for investment obligation are reasonable, we note the language of section 23 is geared towards investment in surveillance and AI technologies. While the obligation for large industry players to undertake collaboration is positive, there is already significant appetite for, and investment in, AI technologies and surveillance mechanisms to counter Class 1A and Class 1B content.

As raised during the consultation briefing, Scarlet Alliance and sex workers and allies across unceded-Australia believe that the missing link between big tech, governance bodies and mechanisms and marginalised communities online is the development of a shared understanding and mitigation of the consequences generated by automatic content-moderation, and development and promotion of 'consent culture' within big tech that facilitates digital access for marginalised communities and provides avenues of redress for unauthorised distribution of adults' intimate images.

The dismissal by the Office of the eSafety Commissioner that unauthorised distribution of intimate images is an 'unenforceable' 'copyright issue' demonstrates a lack of understanding of the harms and safety consequences generated by unauthorised sharing of sensitive digital images. It is clear that collaboration and investment is required to ensure that RES cultivate 'consent culture', and provide clear avenues of redress for users who experience unauthorised distribution of sensitive images.

Recommendation 3: That alongside existing investment requirements, section 23 requires large RES to direct investment towards promoting 'consent culture', facilitating digital access for marginalised communities, and addressing algorithmic discrimination.

Reporting and complaints mechanisms

Are the end-user reporting requirements workable for the relevant service providers? Are there practical barriers to implementation?

While Scarlet Alliance supports the requirements to ensure that users are able to easily report harmful content, we strongly believe that appeals processes against content moderation decisions must be similarly accessible, with requirements for services to respond to appeals.

Sex workers in Australia are frequently the targets for vilification online,[16] and Scarlet Alliance has received anecdotal reports that this often takes the form of malicious reporting of breach of terms of service, with sex workers losing access to reported accounts despite not having committed any breach, and appeals taking months to be reviewed if they are reviewed at all. An obligation to make appeals mechanisms similarly accessible (including an obligation to review requests for appeal) would provide fairness to both people who view harmful content online and people who have been incorrectly categorised as distributing prohibited content.

Recommendation 4: That section 27 of the RES Standard includes an obligation to create mechanisms for end-users and account holders to appeal inaccurate or malicious allegations of distribution of Class 1B content.

Draft Designated Internet Services (DIS) Industry Standard

DIS categories and risk assessment

Are the categories in Table 2 sufficiently clear for designated internet service providers to identify which category they fall within and therefore what obligations apply? What are the benefits and/or challenges of the categories as they are currently proposed?

Our comments on the DIS categories and risk assessment analysis are outlined above in the section on risk assessment. Our recommendation can be found above at **recommendation 1.**

Generative AI in DIS

eSafety is seeking to place requirements on service providers that are best-placed to prevent the use of generative AI features to create and disseminate class 1A and class 1B material. Does the proposal achieve this?

[16] Respect Inc. and Scarlet Alliance, Submission 26 to the Queensland Legal Affairs and Safety Committee on the Criminal Code (Serious Vilification and Hate Crimes) and Other Legislation Amendment Bill 2023 (12 May 2023) 3-4.

While developers must minimise the risk of emerging generative AI technologies being manipulated to create Class 1A and Class 1B content, Scarlet Alliance notes that this risk is one of many presented by unregulated and insufficiently-tested emerging generative AI technologies.

It is vital that regulators consider the privacy and safety risks presented by generative AI beyond the creation of high impact content - as raised in our discussion under the heading Investment requirement for large relevant electronic services above, we believe the Office of the eSafety Commissioner does not have sufficient understanding of the risks generated *for sex workers* from AI-generated 'high impact' content. Before this Standard is implemented, we urge that the Office of the eSafety Commissioner consult with sex workers in Australia to understand the community's privacy and safety concerns in relation to generative AI.

Recommendation 5: That the Office of the eSafety Commissioner conduct consultation with sex workers in Australia in order to understand the implications of 'high impact' generative AI on sex worker privacy and safety.

Investment requirement for large designated internet services

Do you agree with this monthly active user threshold, or are there other thresholds which can be deployed to ensure this obligation is proportionate?

As reflected in our discussion of Investment requirement for large relevant electronic services above, we believe the monthly active user threshold for large DIS is appropriate. However, we believe that this investment must also be directed towards the development and promotion of 'consent culture', digital access and understanding and reducing algorithmic discrimination.

Recommendation 6: That alongside existing investment requirements, section 24 requires large DIS to direct investment towards promoting 'consent culture', facilitating digital access for marginalised communities, and addressing algorithmic discrimination.

Compliance costs

What are your views on the likely compliance costs for service providers and, in particular, the impact of compliance costs on potential new entrants?

As noted in our discussion of Risk assessment above, the pre-assessment of all pornography DIS as Tier 1 essentially generates the highest compliance obligations for all pornography sites and apps. The Office of the eSafety Commissioner is already aware that most pornography in Australia is produced by independent operators, who are often women or LGBTQI+ people.[17] The pre-assessment of these producers as 'high risk' generates significant compliance costs for smaller independent producers, threatening the independence, diversity and more 'ethical' production of

[17] *Roadmap for Age Verification and Complementary Measures to Prevent and Mitigate Harms to Children from Online Pornography* (n 8) 34.

Australian adult-content producers in favour of large international aggregators for whom compliance is less burdensome.

Our recommendation for this section can be found above at **recommendation 1:** That pornography sites and apps are not pre-assessed as 'Tier 1 - High Impact DIS', as there is no evidence that pornography sites as a whole are a higher risk for Class 1A and Class 1B material. Pornography DIS should be subject to the same risk-based self-assessment as other non pre-assessed DIS.

Recommendations

Recommendation 1 (assessment categories): That pornography sites and apps are not pre-assessed as 'Tier 1 - High Impact DIS', as there is no evidence that pornography sites as a whole are a higher risk for Class 1A and Class 1B material. Pornography DIS should be subject to the same risk-based self-assessment as other non pre-assessed DIS.

Recommendation 2 (draft RES Standard 'known' and 'new' content)
A: That the RES Standard include an obligation that services provide users with clear information on the nature and extent of information collected and detection measures used,
B: That the RES Standard clarify (either in the 'example' section following the definition of 'known child sexual abuse material' in section 6, or in the explanatory statement) which non-governmental organisations are deemed to be 'generally recognised as expert or authoritative' in identifying known CSEM, and which body is responsible for making these determinations,
C: That sections 21-22 include specific recognition that services are not expected to break end-to-end encryption,
D: That section 22 includes specific recognition that AI moderation is one of a suite of tools utilised to 'disrupt' and 'deter' unverified Class 1A and Class 1B content, and include examples (either in the Standard or explanatory statement) of non-AI accountability measures.

Recommendation 3 (investment requirement for large RES): That alongside existing investment requirements, section 23 requires large RES to direct investment towards promoting 'consent culture', facilitating digital access for marginalised communities, and addressing algorithmic discrimination.

Recommendation 4 (RES reporting and complaints mechanisms): That section 27 of the RES Standard includes an obligation to create mechanisms for end-users and account holders to appeal inaccurate or malicious allegations of distribution of Class 1B content.

Recommendation 5 (Generative AI DIS): That the Office of the eSafety Commissioner conduct consultation with sex workers in Australia in order to understand the implications of 'high impact' generative AI on sex worker privacy and safety.

Recommendation 6 (investment requirement for large DIS): That alongside existing investment requirements, section 24 requires large DIS to direct investment towards promoting 'consent culture', facilitating digital access for marginalised communities, and addressing algorithmic discrimination.