



ANNEXURE A TO THE ABSTRACT

Annexure A

Background paper, not the primary submission

Online Safety Amendment (Digital Duty of Care) Bill 2026

Supporting paper to the four-page abstract of 7 October 2026.

LODGED WITH

Department of Infrastructure, Transport, Regional
Development, Communications, Sport and the
Arts

DATE

7 October 2026

ORCID

0009-0007-9715-0951

AUTHOR

Viveka Mohan Das
Founder, AISEARCH Global
ABN 13 490 201 041.
Personal capacity. Not a submission on behalf of
any client.

CONSULTATION

Exposure draft released 8 September 2026.
Closed 12pm, 22 September 2026.

CONTACT

viveka@aisearch.global
aisearch.global

Author: Viveka Mohan Das
Location: Camden LGA, NSW (Dharawal Country)
Date: 7 October 2026
ORCID: 0009-0007-9715-0951
Capacity: Founder, AISearch Global (ABN 13 490 201 041). Personal capacity. Not a submission on behalf of any client.
Email: viveka@aisearch.global

ACKNOWLEDGMENT OF COUNTRY. *Written on Dharawal Country in the Camden LGA. Digital harms do not fall evenly. First Nations communities should be involved in co-design of harm categories, not only consulted after a ministerial instrument is settled.*

Not legal advice. Not peer reviewed. Claim labels: **[Verified]** primary source checked; **[Reported]** secondary source; **[Proposed]** recommendation. This annexure was revised on 7 October 2026 to remove overstated penalty figures, to separate the OpenAI incidents from online-safety harms, and to drop comparative examples that do not bear on this Bill.

STATUS Background only. The late-evidence request is confined to the abstract. Points in this annexure on tiering, research access, transparency, First Nations co-design, children and mitigation measurement do not depend on post-deadline evidence.

PENALTY FIGURE Section 26B specifies 60,000 penalty units. From 1 July 2026 the penalty unit is \$364. That is \$21.84 million for an individual. For a body corporate the Regulatory Powers (Standard Provisions) Act 2014 multiplier produces a maximum of \$109.2 million. Earlier drafts of this paper said \$19.8 million. That figure was wrong.

Positions, Corrected

- 01 Answer engines need an express classification.** Section 25A(1)(j) is conjunctive (generate and share). One-to-one answer engines may already fall under designated internet services or search engines. The Bill should say which, instead of leaving coverage to inference.
- 02 Tier by harm profile, not headcount alone.** "Reasonably practicable" in the duty is an after-the-fact standard. Small, invite-only, no-public-content services need a published lower tier. The Minister's exemption power in s.25A(3) is not a planning tool.
- 03 Risk assessments should scale.** An annual written assessment retained for six years fits large services. A lower tier can meet the duty by a biennial declaration against a Commissioner template, still retained and producible on request.
- 04 Research access should not be university employment only.** Section 205B excludes practitioner researchers. A second pathway, with ethics, conflict disclosure and a public-interest test, matches the vetted-researcher logic of Digital Services Act Article 40 more closely than a university monopoly does.
- 05 Transparency for answer engines is citation method, not model weights.** Part 12A's list is built for content moderation. For answer engines the missing item is how sources are selected, ranked and attributed on high-stakes topics.
- 06 First Nations harm categories need co-design** before the Minister adds categories under ss.25C and 25D.

- 07 Complaint duties need a published proportionality test** so micro-services are not prescribed by default.
-
- 08 Do not freeze the eSafety architecture** against the Rickard review's longer institutional options. Review the duty within three years of commencement.
-
- 09 Eighteen-month runway** before the s.26B penalty is enforceable against services below a published harm threshold, with templates issued in that window.
-
- 10 Measure whether mitigations work.** Identification plus a written step is not a duty of care if the next assessment need not say whether harm fell.
-
- 11 Incident notification** within 72 hours of awareness for material online safety incidents, aligned with SOCI Act and Notifiable Data Breaches reporting where both apply. See the abstract. This annexure does not repeat the OpenAI chronology.
-
- 12 Children's best interests as a primary consideration** in risk assessments for services reasonably likely to be accessed by children, not only a banned-content list.
-

1 Standing and Interests

The author advises Australian organisations on how they appear in AI-generated answers. Income therefore depends in part on the transparency of services this Bill would regulate. That interest is disclosed. It is not a reason to treat non-citation of a business as an online-safety harm. The safety claim in this paper is limited to inaccurate or unsafe answers on health, financial, legal and emergency topics, and to unintended tool-use.

The author also convenes The Wattle Quorum, an invite-only practitioner network. As at 7 October 2026 it has no enrolled members. It is not relied on as evidence of platform scale. If it operates as an online service it would be a small, low-harm candidate for the lower tier proposed below. The claim that it is "the only" Australian network of its kind is withdrawn.

No government, platform or industry funding was received for this paper. TCSG (Total Cost per Successful Goal) is the author's unvalidated framework and is not proposed as a statutory metric.

2 Duty, Scope and Proportionality

Schedule 2 would require a person responsible for an online service to ensure, so far as is reasonably practicable, a safe online environment. Failure attracts 60,000 penalty units: \$21.84 million for an individual and up to \$109.2 million for a body corporate. **[Verified: s.26B; penalty unit value from 1 July 2026; corporate multiplier under the Regulatory Powers (Standard Provisions) Act 2014.]** The figure is existential for a small operator even if never imposed.

"Online service" in s.25A(1) covers the existing OSA categories and adds services that allow users to generate AI material and share it. The Minister may exempt services that pose little risk or are used minimally in Australia. An exemption by legislative instrument is not a compliance path a new service can plan around.

The Rickard statutory review is dated October 2024 and was released in February 2025. It recommended duties calibrated to scale and risk, and that obligations fit for a major platform are not fit for a small community forum. **[Reported: Report of the Statutory Review of the Online Safety Act 2021, Delia Rickard, October 2024, released February 2025.]** The exposure draft does not create tiers. "Reasonably practicable" weighs likelihood, degree of harm, knowledge, cost and privacy after the event. It does not tell a one-person service what to file.

[Proposed] Publish a harm-profile tier, in the Bill or in Commissioner guidance. Invite-only services with no public user-generated content and no reasonable likelihood of seriously harmful material should meet the assessment duty by a biennial declaration on a Commissioner template, retained for six years and producible within 30 days. Re-tier when the service changes in a way that alters the harm profile. A user-count threshold alone is the wrong test. A dual test – reach and harm class – is closer to the Dutch platform model and to the Digital Services Act distinction between ordinary services and very large online platforms.

3 Answer Engines

Section 25A(1)(j) applies where users can generate material by AI and share it by a covered service. ChatGPT-style products may still be designated internet services. Google AI Overviews may travel with the search-engine category. That is an argument about the existing list, not a finding that the conjunctive limb covers them.

Ofcom's reading of the UK Online Safety Act is the cleanest comparison. A chatbot is outside the Act if interaction is only with the bot, the bot does not search multiple sites or databases, and it cannot generate pornographic content. Sharing with other users brings it in. The Crime and Policing Act 2026 leaves the UK government a power to bring further chatbot functions into scope, which is itself evidence that the 2023 Act did not settle the question. **[Reported: Ofcom, "AI chatbots and online regulation"; UK Online Safety Act 2023 service definitions.]**

The harm to regulate is a fabricated or materially incomplete answer on a medicine, a payment, a court date or an emergency number. Ranking methodology for commercial queries is a separate transparency issue and should not be smuggled in as safety. **[Proposed]** Either confirm in explanatory material that answer engines and search-integrated generative features are covered, or add a category that does not depend on a sharing function. In Part 12A, require a service-level description of source selection, ranking and attribution for high-stakes topics. Not model weights. Not training data.

4 Risk Assessments and the Measurement Gap

Section 26A requires a written assessment at least annually, before material changes, retained for six years, and available to the Commissioner within 30 days. The Commissioner's call-in power should stay. The default frequency should not be identical for every service.

The deeper gap is that the assessment can record a mitigation without saying whether it worked. The Rickard review recommended reporting on how successful mitigation has been. The exposure draft does not.

[Proposed] For each material risk, record a baseline, the intended effect, and – at the next assessment – whether the indicator moved. A failed mitigation at a higher tier should trigger an earlier reassessment. This does not prescribe a metric.

5 Researcher Access

Section 205B defines an approved researcher as a person employed by an Australian university, approved under the rules, and meeting any further rule requirements. Online-safety research must relate to the Act, be approved by a university human research ethics committee, and be prescribed as in the public interest.

[Verified: s.205B, exposure draft.]

University employment is a proxy for ethics oversight. It is a poor proxy. It excludes practitioners who see answer-engine failures in health and small-business information, and it excludes networks that can publish a charter, a conflict register and an ethics process without a university payroll. Digital Services Act Article 40

vets researchers by independence, funding transparency and security, not by employer class. **[Reported: Regulation (EU) 2022/2065, Art. 40.]**

[Proposed] Keep the university path. Add an accredited-network path: published ethics standard, public conflict register, research related to the Act, a proposal lodged with the Commissioner, and no commercial affiliation with the platform whose data is sought. Decide applications within 60 days. Renew every three years. Do not treat accreditation as a right to model weights.

6 Transparency

Sections 192C and 192M let the Commissioner require safety transparency reports and publication of specified operational information. The listed topics – moderation volumes, complaints, suspensions, user counts – fit social media. They do not describe an answer engine's citation layer.

Automated systems without a disclosure of method have caused serious harm in other settings. The Dutch childcare-benefits inquiry (2021) and the UK Post Office Horizon judgments are cautionary, not analogues of this Bill. They are not relied on as evidence of what section 192M currently covers. **[Proposed]** Add, for services that generate answers to user queries, a duty to publish the design principles for selecting, ranking and attributing sources, at a level an independent reviewer can use to assess systematic error on high-stakes topics.

7 First Nations Digital Safety

Sections 25B and 25C define a safe environment and seriously harmful material. Identity fraud aimed at cultural heritage, disclosure of restricted cultural material, and racial harassment that sits under a general "serious harm" test are not obviously inside the thirteen categories. **[Proposed]** Before the Minister adds a category under s.25C(2) or makes a child-harm determination under s.25D(2), require co-design with First Nations peak bodies, not only Commissioner advice.

8 Implementation

A \$109.2 million corporate maximum will chill small services even if enforcement is aimed at large ones. The EU Digital Services Act staggered very-large-platform duties (August 2023) and other duties (February 2024), with guidance. Australia's draft is more compressed. **[Proposed]** Eighteen months after Royal Assent before s.26B is enforceable against services below a published harm threshold. In that window, publish the tier guide, the template assessment, and the declaration path. Review the duty within three years, including fit with any later institutional reform of the kind the Rickard review left open.

9 Children

The Bill protects children through harmful-material categories and design limits for under-16s on social media. That is not the same as making the best interests of the child a primary consideration in design and in the s.26A assessment. The Rickard review recommended the stronger standard. The UN Convention on the Rights of the Child and the UK Age Appropriate Design Code already use it. **[Proposed]** For a service reasonably likely to be accessed by children, the assessment must treat children's best interests as a primary consideration. If another interest overrides them, record why. This is a governing standard for the existing assessment, not a second assessment.

10 Incident Notification — Cross-Reference

The chronology, the evidentiary limits, and the regulatory-home point are in the abstract. In short: OpenAI says it became aware in mid-August 2026 of June access to Australian government sites, and emailed a public Services Australia inbox on 10 September. The Prime Minister said no individual's medical data was accessed. Related activity involved Victorian health, NSW crime statistics and the Australian Institute of Health and Welfare. These are cyber incidents. They illustrate disclosure delay. They do not by themselves prove an online-safety incident under sections 25B and 25C.

Article 73 of the EU AI Act uses 15 days, 10 days where a person has died, and 2 days for widespread infringement or serious irreversible disruption of critical infrastructure. Article 55(1)(c) covers general-purpose models with systemic risk. A 72-hour online-safety notice is stricter than the EU default and should be stated as such. It should not displace the Security of Critical Infrastructure Act or the Privacy Act Part IIIC clock. MediSecure shows a notice duty can be met while individual notification still fails. It is not the same pattern as a late email to a public inbox.

[Proposed] Preliminary notice to eSafety within 72 hours of awareness of a material online safety incident; fuller report in 30 days; material delay and knowingly incomplete notice as separate penalty grounds; cross-referral where ASD or the OAIC is the lead.

11 Agentic Systems

The Bill has no definition of a system that takes external actions without a human instruction at each step, no tool-use assessment, and no log. OpenAI told Parliament its models took actions it did not intend. That sentence is not an accountability scheme. No UK or US statute currently supplies a full agentic duty.

[Proposed] Define the class narrowly. Require a pre-deployment tool-use assessment, an audit log of material actions, and 72-hour notice of an unintended action affecting an Australian user, personal information, or a third-party system, with cyber incidents notified first through the cyber channel.

Attachment — Interests

AIsearch Global (ABN 13 490 201 041) advises on visibility in AI-generated answers. That commercial interest is relevant to transparency recommendations and is disclosed.

The Wattle Quorum is an invite-only network with no enrolled members at the date of this paper. It is not cited as a representative platform.

AI6 Trust Stack and **Mindful Machines Press** are the author's projects. They are not proposed as regulatory standards.

No funding from government, platforms or industry was received for this paper. Positions and recommendations are the author's. The exposure draft and the Rickard report were read as primary sources. OpenAI's 29 September 2026 post and the Prime Minister's 24 September 2026 remarks were used for the incident account.

AI tools assisted research and drafting. A claim is marked [Verified] only where the author checked a primary source. The author is responsible for the content.

[Verified] checked against the cited primary text. **[Reported]** secondary source. **[Proposed]** recommendation. Revised 7 October 2026. This annexure is background to the four-page abstract.