ALTERNATIVE POLICY · EXPOSURE DRAFT RESPONSE

Online Safety Without Political Censorship

A replacement policy for the Online Safety Amendment (Digital Duty of Care) Bill 2026. Keep the duty to stop crime and predation. Kill the power to police ideas.

Status	Labor exposure draft released 8 September 2026. Consultation closes 12pm, 22 September 2026. Not yet introduced.
Position	Oppose the draft in its present form. Do not amend it into existence. Withdraw it and legislate a narrow Digital Harm Prevention Act.
Core claim	Child protection is a legitimate end. An expandable ministerial catalogue of "harm," enforced by $109 million penalties, is a speech law wearing a safety vest.

1. Executive verdict

The Albanese Government has released an exposure draft that pretends to be a product-safety law for the internet. Parts of it are defensible: child-sexual exploitation material, non-consensual intimate images, nudify tools used against children, suicide-instruction pipelines, and a genuine adult right to switch off an engagement algorithm. Those things can be written in hard language and enforced hard.

The rest of the architecture is the problem, and it is not accidental. Proposed section 26(1) requires every person responsible for an "online service" to "ensure, so far as is reasonably practicable, a safe online environment." Section 25A puts ISPs, social media, messaging, websites, hosting, search, app stores, hardware and AI-generation services inside that duty. Sections 25C(2), 25D(2) and 25G(2) let the Communications Minister add new adult harms, new child harms and new "negative behavioural" design features by legislative instrument. Breach of the duty is 60,000 penalty units (s 26B). That combination does not require a minister to delete a post. It requires platforms to delete first and argue never.

Labor will say this is "not government control" and "My Feed, My Way." Feed choice is not locked in as a statutory right. Section 26(4) lets the Minister require "specified user empowerment tools" by instrument, "without limitation" as to how they are specified. The phrase "hostile attitudes towards women or gender equality" is not a slogan on a lectern. It is proposed s 25D(1)(c). The words "freedom of expression" do not appear in the draft.

This policy therefore does three things. It pulls the draft apart on its own terms. It states the legal and democratic limits Labor is skating over. It sets out replacement legislation that is harsher on crime and narrower on speech.

2. What the exposure draft actually says

Do not debate the press conference. Debate the 8 September 2026 exposure draft. There is no explanatory memorandum. Simplified outlines throughout the draft still "to be drafted."

2.1 Scope — s 25A

An "online service" is an internet carriage service, social media service, relevant electronic service, designated internet service, hosting service, internet search engine service, app distribution service, a service that manufactures or installs equipment for those services, and a service that lets users generate material by AI and share it. A person "responsible" for the service is the provider or anyone in a position of day-to-day

read control. The Minister may exempt a service or class by instrument if satisfied it poses "little risk" or is "used minimally in Australia" (s 25A(3)). The same office that writes the duty decides who is out.

That is how you turn a child-safety announcement into a standing power over the information environment.

2.2 The duty — ss 25B, 26, 26A, 26B

Section 26(1): a person responsible for an online service "must ensure, so far as is reasonably practicable, a safe online environment." That is the digital duty of care. To comply, the person must also manage design features, conduct written risk assessments under s 26A, and take "effective measures" to address them. Failure is a civil penalty of 60,000 penalty units (s 26B). The same figure attaches to ignoring a remedial direction (s 26D).

"Safe online environment" (s 25B) means adults protected from "seriously harmful material and conduct," children protected from "harmful to children" material and from design-feature harms, and — on social media — design features with "negative behavioural impacts" switched off for Australians under 16. The Minister may declare that a service is or is not social media for that last limb (s 25B(2)).

"Reasonably practicable" (s 25H) weighs likelihood, degree of harm, knowledge, cost and privacy. It does not weigh freedom of expression or political communication. The only speech-adjacent carve-out is s 26(7): no action is required on lawful private communications solely between consenting adults. Public posts, ranking, search, group spaces and AI outputs stay in.

2.3 The two lists — ss 25C and 25D

Section 25C ("seriously harmful material and conduct") covers ground a narrow bill should keep: child-sexual exploitation and abuse, grooming, sexual violence, extreme violence, rape and death threats, menacing or harassment intended to cause serious harm, suicide and self-harm instruction, sadistic exploitation, animal cruelty, listed terrorist organisations, praising terrorist acts, illicit-drug instruction, abhorrent violent conduct.

Then the list breaks. Section 25C(1)(j) takes in material that "encourages, promotes, incites or provides instruction for the commission of a criminal offence" — any offence, not only serious indictable crime. Section 25C(1)(m) plus 25C(2) lets the Minister add "any other material or conduct" she is satisfied "may cause serious harm."

Section 25D ("harmful to children") is the political list:

- (a) pornography;
- (b) material or conduct that encourages, promotes or provides instruction for disordered eating;
- (c) material or conduct that encourages or promotes hostile attitudes towards women or gender equality;
- (d) material or conduct that glorifies crime or encourages dangerous stunts or harmful practices;
- (e) abuse, harassment or bullying;
- (f)(i) any other material or conduct that could inflict serious harm to a child;
- (f)(ii) whatever the Minister adds under 25D(2).

Paragraph (c) is viewpoint language. It is not "incitement to violence against women." Paragraph (f)(i) is a residual so wide it makes the rest of the list decorative: anything that "could" inflict serious harm. Eating-disorder instruction is a real harm. "Hostile attitudes towards gender equality" is a factional phrase with a penalty behind it.

2.4 Six instrument powers — the real engine

This is the cluster that should kill the bill. The draft gives the Minister legislative-instrument powers to:

- s 25C(2) — add new "seriously harmful" material or conduct if satisfied it "may cause serious harm";
- s 25D(2) — add new "harmful to children" material or conduct if satisfied it "may cause harm to children";
- s 25G(2) — deem extra design features to have "negative behavioural impacts";
- s 25A(3) — exempt services or classes from the Part;
- s 25B(2) — declare that a service is, or is not, social media for the under-16 feature ban;
- s 26(4) — require specified services to provide specified "user empowerment tools," "without limitation" as to how they are specified.

Section 27B requires the Minister to seek eSafety's advice first. Advice is not a veto. The instruments are disallowable. That is the entire safeguard. The words "freedom of expression" do not appear in the bill. A Parliament that votes for this is not voting for a list of harms. It is voting for a machine that can grow new harms without a new Act.

2.5 Enforcement — ss 26B–26D and Schedule 1

Duty breach and remedial-direction breach: 60,000 penalty units. For a body corporate that is the figure the government has advertised at up to $109.2 million. Transparency, complaints, data-access, link-deletion and nudify-notice failures sit at 6,000 units. Schedule 1, which starts the day after Royal Assent, shortens existing takedown clocks from 48 hours to 24, lets eSafety waive the "complain to the platform first" step, treats screenshots and cropped copies as "reposted," and gives search engines and app stores 24 hours to drop nudify apps and links (ss 86B–86C). Schedule 1 is mostly keepable. Schedule 2 is not.

eSafety is an independent statutory office. Independence from the daily press conference is not independence from incentives. An office measured by "safety outcomes," armed with document compulsion, annual risk papers and nine-figure penalties, will prefer over-removal. That is rational. It is also how lawful speech dies without a criminal charge.

2.6 The decoy — s 26(4), not a statutory opt-out

"My Feed, My Way" is a press release. In the draft, user controls are whatever tools the Minister specifies by instrument under s 26(4). That power can create a real chronological opt-out. It can also create something else. Separate feed choice into a short standalone schedule with the right written in the Act. Do not leave it on a regulation hook glued to the duty.

3. Pulling the legislation apart

Treat the exposure draft as a set of design choices. Each choice has a cost. Most of the costs are being hidden behind children.

Defect 1 — "Harm" is a political solvent

A duty of care works in workplace law because the harm (a crushed hand, a fall from a roof) can be described. Online "harm" in this draft includes s 25D(1)(c) "hostile attitudes towards women or gender equality," s 25D(1)(f)(i) anything that "could inflict serious harm to a child," and two ministerial residuals that turn on what the Minister is "satisfied may" cause harm. Section 25E even says harm can be caused by material or conduct "whether or not engaged in or experienced by a user of an online service."

Vague harm fails the first discipline of a free society: a citizen and a company must be able to know the rule before the fine. It also fails ICCPR Article 19(3)'s demand that restrictions be "provided by law" — meaning precise, accessible, and not a moving target.

Defect 2 — The duty privatises censorship

Labor will never have to ban a viewpoint if Facebook, Google, TikTok, OpenAI and every forum host will suppress it to stay on the safe side of a $109 million risk. That is not a side-effect. That is the enforcement model. The state sets an elastic standard. Private firms, terrified of systemic findings, apply a tighter standard. Parliament keeps its hands clean. The speech still disappears.

This is why comparisons with seatbelts and toys are dishonest. A seatbelt standard does not require Toyota to guess next year's definition of a crash. This duty requires platforms to guess next year's definition of a bad idea.

Defect 3 — Section 25D(1)(c) is viewpoint discrimination

The draft's words are "encourages or promotes hostile attitudes towards women or gender equality." If that means incitement to violence against women, use the criminal law that already exists. As written it is a political filter. The same clause will be read, by a different government, against radical gender theory, against religious teaching on sex, or against whoever lost the last culture war. A serious legislature does not enact a weapon and trust its own side to hold it forever.

Defect 4 — Sections 25C(2), 25D(2) and 25G(2) make the Act a husk

If harm categories and design-feature lists can be added by instrument, the Parliament is not settling the boundary of lawful speech. It is outsourcing it. Section 27B (ask eSafety first) and Senate disallowance are weak checks. The bill does not even recite freedom of expression. Compatibility statements, when they eventually appear in an EM that does not yet exist, will be written by the same department that wants the power.

Any replacement statute must put every restricted category in the Act. No new speech-restrictive category by regulation. Full stop.

Defect 5 — Systems and design duties swallow political communication

Once you regulate "algorithmic content recommendation or generation systems" under a general safety duty, you are no longer only targeting illegal files. You are targeting ranking, discovery, amplification and, in AI systems, the answers a model is willing to give. That is the public square of 2026. A duty to make those systems "safe" against undefined psychosocial harm is a duty to flatten controversy.

Australia's implied freedom of political communication is not a personal First Amendment. It is a limit on laws that burden communication about government and political matters without being reasonably appropriate and adapted to a legitimate end. A systems duty that predictably suppresses contested political speech about sex, crime, immigration, medicine and the government itself is exactly the kind of burden the High Court is for.

Defect 6 — The ICCPR argument Labor hopes you will botch

Article 19(1) of the ICCPR — the right to hold opinions — cannot be restricted. Article 19(2) covers seeking, receiving and imparting information and ideas of all kinds, by any media, including ideas that offend. Article

19(3) allows limits only if they are provided by law, aimed at a listed purpose (rights of others, national security, public order, public health or morals), and necessary and proportionate.

Australia ratified the Covenant in 1980. It has not enacted a federal Human Rights Act. So Article 19 is a binding international obligation and a scrutiny standard, not a free-standing federal cause of action. That does not make it decorative. It is the test Labor itself recites in compatibility statements. On that test this draft is weak where it is vague, weaker where a minister can enlarge it, and weakest where a nine-figure penalty produces collateral suppression of lawful political speech.

Children's rights pull the other way. They justify targeted rules. They do not justify a general power to treat adult political argument as a child-safety defect and then apply the same machinery across the open internet.

Defect 7 — Mission creep from a bill Labor could not pass

The misinformation and disinformation project died because too many people noticed that "false and harmful" is a political weapon. The Digital Duty of Care is the successor technology: drop "false," keep "harm," move from content takedown to systems liability, and dare opponents to look as if they want teenagers on pornography and suicide forums. That is competent politics. It is contemptuous legislating.

Defect 8 — eSafety is being built into a standing speech office

Give a regulator removal notices, document compulsion, faster clocks, a duty to assess "systems," and penalties that matter to global firms, and you have built an institution that will expand. Institutions do. The personal virtue of any particular Commissioner is irrelevant. The incentive is the policy. A replacement statute should shrink the office back to illegal and clearly defined child-harm material, with merits review that is real, and with an express political-communication carve-out a court can use.

Defect 9 — Surveillance travels with the duty

A duty to prevent foreseeable harm at systems level pushes firms toward more identity, more age assurance, more logging, more scanning. Some of that is justified for Class 1 material. Applied to "hostile ideas" and self-esteem features it becomes a reason to end practical anonymity in ordinary political speech. Anonymity is not a licence to commit crimes. It is often the only reason a nurse, teacher, soldier or public servant will criticise a government program. A safety duty that casually burns that is not safety. It is discipline.

Defect 10 — Fourteen days of "consultation" on a speech architecture

The exposure draft landed on 8 September. Feedback closes at noon on 22 September, capped at five pages, with no guarantee of publication. That is not how a liberal parliament treats a law that can restructure the public sphere. It is how a government treats a drafting exercise it has already decided to run.

4. What to kill, what to keep

KILL	KEEP — AND TIGHTEN
s 26(1) general duty to "ensure a safe online environment." ss 25C(2), 25D(2), 25G(2) ministerial add-ons. s 25D(1)(c) and (f)(i).	Duty to take reasonable steps to prevent Class 1 material and other already-criminal content being distributed on the service.
"Hostile ideas about women and gender equality" as a harm category.	Existing criminal offences for incitement, threats, stalking, image-based abuse and child-sexual offences — enforced, not re-labelled.

Ministerial or eSafety power to designate additional harms by instrument.	Every restricted category written into the Act. No speech-restrictive add-ons by regulation.
Open-ended "mental wellbeing" / "psychosocial harm" / "public safety" residual categories.	Tight child-specific rules on pornography, suicide and eating-disorder instruction material, using classification language not political slogans.
$109 million penalties for failing a paper risk-assessment against vague harms.	Serious penalties for systemic failure on illegal content and child-sexual material. Graduated, court-supervised, with a defence of reasonable systems aimed at defined categories.
Systems duty over ranking and model outputs aimed at "hostile ideas."	Adult user control: a real, durable, default-respecting opt-out of engagement ranking. No government-chosen replacement ranking.
Using child safety as a wrapper for adult political speech.	Removal powers for nudify tools and non-consensual intimate imagery, including of children. Fast clocks. Real review.

5. Replacement policy — the Digital Harm Prevention Act

Do not try to "improve" Labor's duty by adding a recital about free speech. Recitals do not restrain a $109 million incentive. Replace the architecture.

Object of the Act

1. Prevent the distribution of material that is already illegal under Commonwealth law, with particular force against child-sexual exploitation, pro-terror material, and image-based abuse.
2. Protect children from a closed list of age-restricted harms defined in the statute in classification language.
3. Give adult users control over engagement ranking on large social media services.
4. Preserve freedom of political communication and Australia's Article 19 obligations by forbidding viewpoint categories and ministerial expansion of speech restrictions.

Part A — Illegal material duty (harsh, narrow)

Every service within the existing Online Safety Act coverage that allows users to upload, generate or recommend content must take reasonable steps to prevent the service being used to distribute:

- Class 1 material, including child-sexual exploitation material;
- pro-terror material as already defined;
- non-consensual intimate images, including synthetic "nudify" output;
- material whose creation or distribution is already a criminal offence (threats to kill, stalking, doxxing where criminal, cyberflashing).

Reasonable steps means functioning hash-matching and known-CSAM disruption, rapid action on valid notices, and no wilful blindness. It does not mean a duty to pre-empt lawful opinions. Failure that is systemic — not a single missed item — attracts serious civil penalties, imposed by a court, with a due-diligence defence tied to the statutory list, not to a minister's later mood.

Part B — Closed child-protection list

For users the provider knows or reasonably ought to know are under 18, the provider must take reasonable steps to prevent the service recommending or, where technically feasible, supplying:

- X18+ and equivalent pornography;
- material that instructs in suicide or serious self-harm;

- material that instructs in the methods of an eating disorder (instruction, not discussion, memoir, news or recovery).

Glorification of criminal offending by children can be added only if defined as material that instructs in or directly encourages the commission of an indictable offence, not "edgy crime content" in the abstract.

What is forbidden on this list: "hostile ideas," "misogyny" as a mood, "gender equality," "mental wellbeing," "self-esteem features," and "social cohesion." Those phrases do not enter the Act. If a future Parliament wants them, it must amend the Act in public, on the record, and wear the argument.

Part C — Adult feed control, without a state algorithm

Large social media services must offer every user 16 and over a durable control: personalised engagement ranking, or a reverse-chronological / follow-only feed. The choice must persist. It must not reset after an "update." The service may not dark-pattern the user back into the engagement feed. The Commonwealth must not prescribe a replacement ranking, a "public interest feed," or a list of favoured sources. That would be the Alborithm. This policy rejects it.

Part D — Speech lock

The Act must contain operative provisions, not recitals:

- Nothing in the Act authorises a requirement to remove, suppress, demonetise, down-rank or refuse to generate material solely because it expresses a political, religious, scientific or moral opinion.
- The Act does not apply to the extent it would infringe the implied freedom of political communication (retain and strengthen the existing s 233 logic).
- No rule, standard, code or notice may create a new category of restricted material. Categories live in the Act.
- eSafety notices must specify the statutory category relied on. A notice that cannot name a category is void.
- A service that removes material under government pressure outside those categories must log the request. An annual public report of government takedown and "voluntary" moderation requests is mandatory.

Part E — Regulator, cut down to size

- eSafety retains teeth on Class 1, image-based abuse, child cyberbullying as presently defined with an objective threshold, and the closed child list.
- Document-compulsion stays for suspected systemic failure on illegal material. It does not become a standing right to inspect ranking models for "hostile ideas."
- Merits review of removal and remedial directions must be real and reasonably fast. A regulator that cannot be reversed will not be careful.
- Civil penalties for the illegal-material duty may be high. Penalties for paperwork failures on undefined "safety culture" must not.
- The Commissioner remains independent of daily ministerial direction on individual cases. The Minister loses the power to grow the harm list.

Part F — What this policy will not do

- It will not ban adults from seeing lawful pornography.
- It will not force identity display for ordinary political speech.
- It will not revive the misinformation bill under another name.

- It will not let a minister declare a newspaper column, a sermon, a feminist essay or a men's-rights rant a "harm category."
- It will not pretend that Silicon Valley is innocent. Reckless product design aimed at children can be attacked with the closed list and with consumer law. It does not require a speech solvent.

6. Legislative instructions — how to vote

1. Vote against introduction of the Online Safety Amendment (Digital Duty of Care) Bill 2026 in the form described in the 8 September exposure materials.
2. If the government proceeds, circulate amendments that strip ministerial harm-designation, delete "hostile ideas" and residual psychosocial categories, lock categories into the Act, insert the speech lock in Part D, and detach feed-choice into a short standalone schedule.
3. Do not accept a "we will fix it in codes" deal. Codes written under a vague duty reproduce the duty.
4. Move for a proper inquiry with published submissions and more than ten business days. A speech architecture does not get a fortnight and a five-page cap.
5. If the government refuses to narrow the bill, vote it down. A bad duty with a free-speech recital is still a bad duty.

7. Answers to the predictable smear

"You want children on pornography and suicide forums."

No. This policy is more specific, and therefore more enforceable, on those exact harms. Labor's draft is weaker where it is vague, because platforms will spend their compliance budget on political risk and call it safety.

"Other democracies do duty of care."

The UK and EU models are not holy writ, and both are already producing over-removal. Copying a foreign error is not leadership. If those statutes contain viewpoint phrases and ministerial add-ons, do not import the infection.

"Freedom of expression is not absolute."

Correct. Article 19(3) and the implied freedom both allow limits. They require precision, a legitimate end, and proportionality. "Hostile ideas" plus a minister plus $109 million is not that test. It is an evasion of it.

"eSafety is independent."

Independence from a party room is not independence from institutional incentives. Give any office a systems duty over "safety" and it will discover more unsafety. That is what offices do.

"This is just like workplace health and safety."

A factory does not have a constitutional function of political communication. A public square does. Analogies that erase that fact are propaganda.

8. Immediate actions

1. Lodge a submission to digitaldutyofcare@communications.gov.au before 12pm on 22 September 2026. Keep it inside five pages if that remains the cap. Put the kill-list and the speech lock on page one.
2. Publish the submission. Labor has said it may not publish yours.

3. Demand the government table the exposure draft's human-rights compatibility analysis against ICCPR Article 19 and the implied freedom, in detail, not as a paragraph.
4. Senators should prepare a disallowance posture now for any later instrument that adds a harm category. Do not wait until the instrument is buried in a sitting week.
5. Attach feed-choice to a short private members' bill so it cannot be held hostage to the duty.

9. Closing

Child-sexual exploitation, image-based abuse, suicide instruction and industrial-scale pornography aimed at minors are not free-speech issues. They are enforcement issues. Labor has every right to be brutal on those.

It does not have the right to smuggle a standing power over lawful political argument into that brutality. "Safe online environment," "hostile ideas," and a minister who can add harms later are not child protection. They are the misinformation project after a costume change.

A serious country writes the banned list in an Act, points the fine at crime, gives adults a switch for the ranking engine, and leaves the argument about sex, gender, God, crime and government where it belongs: in public, where it can be answered.

Withdraw the Digital Duty of Care Bill. Legislate the Digital Harm Prevention Act. Anything else is an admission that the children were the pretext.

Document control

Purpose: public policy position and drafting brief. Primary text: Exposure Draft, Online Safety Amendment (Digital Duty of Care) Bill 2026 (8 September 2026), especially proposed ss 25A–26D, 25C(2), 25D, 25G(2), 26(4), 26(7), 26B, 27B and Schedule 1 ss 86B–86C. No explanatory memorandum has been published. This paper is an opinionated legislative alternative. It is not legal advice. Also: ICCPR Article 19; General Comment No. 34; Online Safety Act 2021 s 233.