

# Digital Duty of Care Exposure Draft Bill 2026

## “Build Your Own Submission” resource

**CONSULTATION CLOSES**
**22 SEPTEMBER 2026 at 12pm**

**NOW UPDATED WITH THE BILL INTRODUCED TO PARLIAMENT ON 15 OCTOBER 2026**

DIGITAL RIGHTS WATCH

WOTTON KEARNEY

# Build Your Own Submission

*Online Safety Amendment (Digital Duty of Care) Bill 2026*

---

This resource contains:

- Information about the Digital Duty of Care Bill and tips to support you with your submission
- Template submission designed for Community Legal Centres, Civil Society and Not-for-Profit Organisations

## CONSULTATION CLOSES: 22 SEPTEMBER 2026 at 12pm

**This document contains general information available at the time of publication. It does not constitute legal advice.**

**UPDATE: 23 OCTOBER 2026**

## THE INTRODUCED BILL

The Bill was introduced to Parliament on 15 October 2026 with an Explanatory Memorandum. This Guide has been kept as originally written and describes the exposure draft, so its section numbers are exposure draft numbers.

We draw your attention to changes that you may be interested in considering for advocacy purposes in Part 1 and Part 2 where relevant under the heading "**UPDATE: WHAT HAS CHANGED IN THE INTRODUCED BILL**". (Note that we haven't worked through or added any updates to Part 3 and Part 4.)

References are labelled "ED" (exposure draft), "Bill" (introduced Bill) and "EM" (Explanatory Memorandum). Many provisions have been renumbered; see the **APPENDIX: CHANGES IN KEY PROVISIONS — EXPOSURE DRAFT COMPARED WITH THE INTRODUCED BILL.**

The Environment and Communications Legislation Committee is taking submissions until **30 October 2026**.

Committee Secretariat contact:
Committee Secretary
Senate Standing Committees on Environment and Communications
PO Box 6100
Parliament House
Canberra ACT 2600
Phone: +61 2 6277 3526
ec.sen@aph.gov.au

# ABOUT THIS RESOURCE

This resource was prepared by Wotton Kearney in partnership with Digital Rights Watch as a practical tool for community legal centres, civil society and not-for-profit organisations to engage with the Australian Government's consultation on the *Online Safety Amendment (Digital Duty of Care) Bill 2026* exposure draft (**the Bill**) which is available here.

**The consultation on the Bill closes at 12pm on 22 September 2026.** Submissions should be sent to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts at digitaldutyofcare@communications.gov.au. See Part 4 of this resource for a template submission.

We have done our best in the short timeframe since the release of the exposure draft to provide enough background information, context and analysis for your organisation to form its own positions, picking the key issues that are most relevant to your work. Our aim is to support you to write a submission in your organisation's voice, drawing on your expertise and clients' lived experience.

> **TIP** Don't let the perfect be the enemy of the good. Say what you support, not just what you'd change. There is currently a lot of confusion about what the Bill means and what impact it will have. Your organisation's real life case studies will be important for progressing informed public consideration of the Duty.
>
> Consider framing your submission around three elements: (1) what you support and want preserved, (2) what you support but want strengthened, and (3) your concerns and what you'd recommend instead.

# TABLE OF CONTENTS

ABOUT THIS RESOURCE ..... 3
TABLE OF CONTENTS ..... 4
PART 1: OVERVIEW OF THE BILL ..... 5
PART 2: DIGITAL DUTY OF CARE – PROS AND CONS & AMBIGUITY (ISSUES FOR YOUR SUBMISSION) ..... 16
PART 3: CASE STUDIES – HOW THE DIGITAL DUTY OF CARE WOULD APPLY ..... 41
PART 4: TEMPLATE SUBMISSION ..... 46
APPENDIX: QUICK REFERENCE — KEY PROVISIONS ..... 49

# PART 1: OVERVIEW OF THE BILL

This Part provides a summary of the Bill. It is designed to give your organisation an understanding of the Bill, and to help you engage with the detailed issue analysis in Part 3 of this resource.

## What the Bill Does

The Bill (available here) amends the *Online Safety Act 2021* (Cth) (**OSA**) (and keeps most of the existing OSA).

It does three main things:

1. **Introduces a Digital Duty of Care** (Schedule 2 of the Bill)
2. **Gives the eSafety Commissioner additional powers** to remove certain content, links and apps (Schedule 1 of the Bill)
3. **Repeals industry codes and industry standards** which currently form part of the OSA (Schedule 3 of the Bill)

**A glossary** of key terms used in the Bill is below at **Part 2.5.**

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

The introduced Bill keeps this three-part structure but adds two new obligations to Schedule 2:

- Services that could expose children to pornography must now check users' ages and restrict access (Bill ss 27D–27F); and
- the eSafety Commissioner can require services to explain how their features and algorithms work (Bill s 28A) (EM pp 76–80).

The Bill also strengthens the existing complaint schemes. The maximum penalty for ignoring a removal notice rises from 500 to 6,000 penalty units (EM p 2), and a platform now has only 24 hours, rather than 48, to deal with a complaint before the Commissioner can step in and order the material taken down (EM p 3).

Finally, a handful of other Acts, including the Privacy Act 1988 and the Criminal Code, are amended so they fit with the new scheme (EM pp 7–8).

# PART 1.1 The Digital Duty of Care

## Overview

The **digital duty of care** requires **a person responsible for an online service** to:

1. Ensure a **safe online environment** so as far as is reasonably practicable
2. **Manage design features** of online services (such as algorithms, endless feeds etc)
3. Conduct regular **risk assessments** and take effective measures to address risks.

## Who owes the duty of care?

The duty of care is owed by '**a person responsible for an online service**'.

**Online services** (section 25A) are defined broadly. The definition is intended to be technology-neutral and have the ability to cover emerging services that don't yet exist. It includes:

- social media services
- search engines
- app stores (including distribution services)
- hosting services
- devices and equipment, and
- services that can generate AI material and share it on another online service.

'Lawful communications in private between consenting adults' (s 26(7)) are not captured, although it's not clear how this would be assessed.

**'A person responsible**' (s 25A(2)) includes a person who provides the service and a person in day-to-day control of it. It seems this is intended to capture both the corporate entity and senior staff with operational responsibility.

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

### The duty

The core duty has been reworded. Instead of ensuring safety "so far as is reasonably practicable", platforms must now "take appropriate action" to ensure a safe online environment (Bill s 26; EM p 51). Section 26A sets out what counts as appropriate, and the cost of an action cannot be used as a reason not to take it (Bill s 26A(3); EM p 67). A single piece of harmful content will not by itself mean a platform has breached the duty, although it may be evidence of a wider, systemic failure (EM p 67).

### Who it covers

Coverage is much the same. The definition of "online service" has moved to section 25B, and it still expressly includes services that use artificial intelligence to generate content (Bill s 25B(1)(h)). State and Territory government agencies are now excluded (Bill s 25A(1); EM p 50).

### Who is responsible

This has been narrowed. The duty now falls on the provider, or on whoever exercises "the greatest degree of day-to-day control" over the service, rather than on anyone in a position to exercise day-to-day control (Bill s 25B(2); EM p 52). This may mean senior staff are no longer covered in the way described above (see section 2.7).

## What does the duty require? 'a safe online environment'

The duty requires online services to ensure 'a safe online environment' (section 25B) so far as is reasonably practicable.

A 'safe online environment' is different for adults and for kids. That is, the Bill requires online services to do more to protect kids, and less to protect adults.

For definitions of terms used in this table, see glossary (part 1.5 below), the appendix to this document, and the Bill itself.

| Adults | Children |
|---|---|
| **HARMFUL MATERIAL AND CONDUCT**<br><br>Protected from **seriously harmful material and conduct (s 25C):**<br>• Child sexual exploitation/abuse<br>• grooming a child<br>• sexual violence<br>• extreme violence or cruelty<br>• threats of violence including rape<br>• menacing or harassing<br>• self harm/suicide material<br>• terrorism<br>• criminal offence<br>• drug use<br>• abhorrent violent conduct<br>• **Other material / conduct decided by the Minister** | **HARMFUL MATERIAL AND CONDUCT**<br><br>Protected from **material and conduct that is harmful to children (s 25D):**<br>• Anything that could inflict serious harm to a child (this is a 'catch-all')<br>• porn<br>• eating disorders<br>• hostile attitudes towards women or gender equality<br>• crime<br>• abuse or bullying<br>• **Anything else determined by Minister (section** 25D(2)). |
| **HARMFUL FEATURES**<br><br>**Those subject to the duty must:**<br>• 'Manage design features appropriately' (*meaning unclear*) **and** | **HARMFUL FEATURES**<br><br>Children must be protected from:<br>• harms associated with certain design features |

| | |
|---|---|
| • Provide 'user empowerment tools' (as required by the Minister under a legislative instrument s 26(4).)<br>  ○ ***(this is relevant to the algorithm opt in/opt out we've been hearing about in the media, which doesn't currently exist under the Bill)*** | • *(only for children under 16 on social media)* design features that have 'negative behavioural impact' (s 25G) being:<br>  ○ a recommender feature (ie algorithm)<br>  ○ a logged-in feature;<br>  ○ an endless-feed feature;<br>  ○ a feedback feed feature;<br>  ○ a time limited feature;<br>  ○ a design feature determined by the Minister. |

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

## Safe online environment

Children under 16 must now "not have access to, or be exposed to" harmful design features, rather than saying those features must not operate (Bill s 25C(1)(c)). The Government also says the duty is not meant to undermine people's privacy, for example, by forcing platforms to break end-to-end encryption. However, that promise appears only in the EM, not in the Bill itself (EM p 53).

## Seriously harmful material

This is the material platforms must protect everyone from (Bill s 25D). It is now a closed list, so the Minister can no longer add new categories (EM pp 54, 57). Four new categories have been added:

- fake nude images,
- sharing or threatening to share intimate images,
- material promoting gender-based violence, and
- violent extremist material (EM pp 55–56).

Harassment is now caught if it is "reasonably likely" to cause serious harm, even if the harasser did not intend that. Other categories have been narrowed. The criminal offence category now covers only serious federal offences carrying 3 years' imprisonment or more (EM pp 49, 56), the illicit drugs category has been removed, and the suicide category now excludes voluntary assisted dying.

## Material harmful to children

These categories are the same as in the exposure draft: pornography, disordered eating, hostile attitudes towards women, glorifying crime or dangerous stunts, and bullying (Bill s 25E). Two things have changed:

- The Minister can no longer add to the list, and
- the general catch-all now covers material that "may cause serious harm to children", rather than material that "could inflict serious harm" (EM pp 57–59).

**Excluded material.** The Bill creates a new category of "excluded material" that falls outside both harm lists, so platforms do not have to act against it (Bill s 25F). It covers:

- professional news;
- material used in court proceedings or by public officials;
- classified films, games and other entertainment;
- research;
- literary, artistic or educational material;
- public health and harm-reduction information; and
- material in the public interest (EM pp 59–61).

### Design features and negative behavioural impacts

For children under 16 on social media, only three types of features are now automatically treated as harmful, or as having "negative behavioural impacts":

- recommender features (algorithms that choose what users see);
- "logged-in features"; and
- any feature the Minister adds (Bill s 25G(1); EM p 62).

Endless scrolling, likes and follower counts, and disappearing content are no longer on the list in their own right. They count only as "logged-in features", meaning they are available only to users who are logged in to an account (Bill s 25G(4); EM p 64). Read receipts are expressly excluded (Bill s 25G(7)). Children of every age are still protected more generally from harm caused by design features (EM p 53).

# PART 1.2 Risk assessments

Online services are required under the Bill to do risk assessments.

A risk assessment is a written assessment of the risk of harm posed by material or conduct done at least annually (section 26B) It must:

- **Identify:**
    - all reasonably foreseeable risks (not just those relevant to the Digital Duty of Care),
    - things that give rise to the risk (Content, features, other systems and processes)
    - people who may be affected by the risk
- **Assess** likelihood and severity of risks
- **Document** risk minimisation measures (implemented and planned)
- **Be provided** to the eSafety Commissioner for regular review and assessment

Risk assessments are to be done at least annually.

Failure to comply with the risk assessment requirements - 60,000 penalty units (s 26B) – up to $109 million for corporations.

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

### Risk assessments

These are now a separate legal obligation, with their own penalty for non-compliance of up to 60,000 penalty units (Bill ss 27, 27B). Assessments must:

- record how effective the mitigation measures are expected to be,
- be done at least annually and before significant changes to the service,
- be kept for 6 years, and
- be given to the Commissioner within 30 days of a request (EM pp 70–71).

The EM says services must consider cumulative harms, such as sustained campaigns of individually minor abuse (EM p 72).

### Other requirements

- **Age checks.** Where children could foreseeably come across pornography or similar material, platforms must check users' ages and restrict access. The penalty is up to 60,000 penalty units (Bill ss 27D–27E; EM pp 76–77).
- **Complaints.** Certain platforms must offer complaint processes that are open equally to all individuals in Australia, whether or not they have an account (Bill s 27G; EM p 78).

# PART 1.3 More powers for eSafety Commissioner

The eSafety Commissioner is given new powers relating to the removal of certain harmful material or places that material is located online under Schedule 1 of the Bill.

## Fake nude apps and websites removal (Part 6A, s 86B)

When it comes to apps or websites that are designed or used predominantly to generate fake nude material, the eSafety Commissioner has new powers to issue notices to (section 86B):

- Remove the app
- Remove links to the app or website
- Remove advertising or information about the app/website

Compliance is required within 24 hours or as otherwise agreed with the Commissioner (s 86C).

Non-compliance attracts penalties of 6,000 penalty units or up to $109 million for corporations.

Online services can seek a review of a notice in the Administrative Reviews Tribunal (s 220(10B)).

## Link deletion notices (Division 3 of the Bill)

The Commissioner can issue link deletion notices to search engines requiring them to cease providing links to certain types of material, specifically:

- cyber-bullying material targeting Australian children or adults (see ss 72A and 92A)
- Intimate images (s 82)

A prerequisite is that a removal notice has already been issued to the hosting service.

Compliance is required within 24 hours or as otherwise agreed with the Commissioner, with penalties of 6,000 penalty units for non-compliance.

### Complaint waiver and reposting

The Bill gives the Commissioner two important procedural flexibilities:

**Complaint waiver**: the Commissioner can waive the usual requirement for a complaint to the service provider where there is no functional complaints mechanism or where making a complaint would risk further harm to the affected person.

**Reposting**: the Commissioner can act on reposted material (even if altered) without requiring a fresh complaint. This addresses the common problem of harmful material being taken down and then re-uploaded in a slightly modified form.

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

### Fake nude material

The power to remove "nudify" apps and websites, which create fake nude images of real people, has been narrowed. In the exposure draft, an app or website could be targeted if it was designed for this purpose or mainly used for it.

The Bill requires both: it must be designed for creating fake nude images and mainly used for that purpose (Bill s 86B). Confusingly, the EM still describes the test as "or" (EM p 25). On the other hand, a removal notice can now also cover copycat apps or websites that are "substantively the same" (EM p 26).

### Link deletion notices

These require search engines to stop showing links to harmful material. They now also cover intimate images (Bill s 82; EM p 33). For the most serious material (Class 1), a link deletion notice can now be issued at the same time as a removal notice, rather than only afterwards (EM p 35). The Commissioner must withdraw a link deletion notice once the material has been taken down, or once the notice is no longer needed (Bill ss 72D, 82C, 92D, 127; EM pp 33–35).

### Complaints and reposted material

The Bill makes three changes:

- **Less time for platforms.** When someone complains to a platform, the platform now has 24 hours, not 48, to deal with it before the Commissioner can step in and order the material to be taken down (EM pp 28–31).
- **Reposted material.** The Commissioner can act on reposted material without a fresh complaint, even if it reappears on a different platform or is posted by a different person (EM p 29). The exposure draft's 6-month time limit on this has been removed.
- **Minimum age penalties.** The exposure draft's proposed increase in penalties for breaching the social media minimum age has been dropped (ED Schedule 2, items 13–15).

# PART 1.4 Enforcement

The Bill gives the eSafety Commissioner a graduated suite of enforcement tools:

- **Formal warnings (s 26C)**: the Commissioner may issue formal warnings where there has been a contravention of the duty.
- **Remedial directions (s 26D)**: the Commissioner may direct a person to take specified action to remedy a contravention. Non-compliance attracts penalties of up to 60,000 penalty units (up to $109 million for corporations).
- **Infringement notices**: for less serious breaches, the Commissioner may issue infringement notices as an alternative to civil penalty proceedings.
- **Civil penalties (s 26B)**: breach of the duty itself attracts civil penalties of up to 60,000 penalty units (up to $109 million for corporations).

The Bill also introduces several transparency and accountability mechanisms:

- **Transparency reports (Part 12A, s 192C)**: services will be required to publish regular transparency reports.
- **Compelled publication of safety information (s 192M)**: the Commissioner may require services to publish specific safety information.
- **Data access for researchers (Part 14, s 205B)**: the Bill establishes a framework for independent researchers to access data held by online services.
- **Sock puppet identity protections (ss 205G–205L)**: protections for researchers and the Commissioner to create accounts for investigative purposes without revealing their identity.
- **Annual statement of monitoring and enforcement priorities (s 166C)**: the Commissioner must publish an annual statement setting out regulatory priorities.

## WHAT HAS CHANGED IN THE INTRODUCED BILL

### Penalties

The maximum penalty for breaching the duty stays at 60,000 penalty units (Bill s 26C). The Commissioner can still issue formal warnings (Bill s 26D). It can also give remedial directions, which are orders telling a platform to fix a problem (now Bill s 26E). However, the penalty for ignoring a remedial direction has dropped tenfold, from 60,000 to 6,000 penalty units (EM pp 68–70).

### Infringement notices

These are fines a platform can pay instead of facing further enforcement action, such as court proceedings. They can be used for all the other new obligations, including remedial directions, but not for a breach of the duty itself. The EM is inconsistent on this point: its general description says infringement notices can be used to enforce the duty (EM p 69), but its detailed notes on the changes leave the duty out (EM pp 81–82). For these obligations, the fine for companies rises tenfold, from 60 to 600 penalty units, while the amount for individuals stays at 12 penalty units (Bill s 163AA; EM pp 82–83).

### New tools for the Commissioner

The Commissioner can now:

- publish formal warnings, so the public knows which platforms have been warned (Bill s 183AA; EM p 89);
- require platforms to explain how their features, systems and algorithms work, or require them to let an expert examine those features and systems and report on them (Bill s 28A; EM pp 79–80); and
- use its information-gathering powers to check whether platforms are complying with any part of the Act, not just particular schemes (EM p 93).

Transparency reporting continues under Bill ss 192B–192K, and the Commissioner can decide what information platforms must publish (Bill s 192L; EM pp 91–92). The sock puppet protections are also kept. These let researchers and the Commissioner use anonymous test accounts to investigate how platforms behave (Bill ss 205H–205K; EM pp 98–99).

# PART 1.5 Glossary

### “Design features” (s 25F)

The Bill defines “design features” to include:

- Recommender systems and algorithmic content curation
- Endless-feed or infinite-scroll mechanisms
- Feedback mechanisms (such as likes, reactions and follower counts)
- Features that require or encourage users to be logged in
- Time-limited features (such as disappearing content)

Note that only children are protected from harms caused by design features of online services.

### “Negative behavioural impacts” (s 25G)

This concept applies specifically to children under 16 on social media. Where design features negatively affect children’s behaviour — for example, by encouraging compulsive use or amplifying harmful content — those features must not operate in respect of those children.

### “Seriously harmful material and conduct” (s 25C)

This term covers defined categories of the most severe online harms, including illegal content, content promoting violence, exploitation material, and other content that poses a serious risk of harm.

Both children and adults are protected from seriously harmful material and conduct.

### “Harmful to children” (s 25D)

This is a broader category than “seriously harmful” material. It includes content involving bullying, self-harm promotion, age-inappropriate material, and other material that may harm children even if it does not meet the higher threshold of “seriously harmful.”

### “Reasonably practicable” (s 25H)

The duty is not absolute. Section 25H sets out the factors for determining what is “reasonably practicable,” balancing:

- The likelihood and degree of harm
- The knowledge (actual or constructive) of the person responsible
- The availability and suitability of ways to mitigate the harm
- The cost of mitigation, including whether the cost would be grossly disproportionate to the risk
- The impact on user privacy

### “User empowerment tool” (s 26(6))

A user empowerment tool lets a user of an online service manage the way design features of the service operate for the user (for example, whether or not they want to have an algorithm determining what content they see).

There is currently no requirement for online service providers to provide user empowerment tools. The Minister has the power to make online service providers provide ‘user empowerment tools’ via a legislative instrument (s 26(4)).

### “Legislative instrument”

The Minister is given a range of powers under the Bill to do things ‘by legislative instrument’. This means that the Minister could make changes to the Bill by tabling a legislative instrument in Parliament. A legislative instrument can be disallowed, or overruled, by Parliament (unless exempted from that, but no exemptions apply under the Bill). For this reason, legislative instruments under this Bill may also be referred to as ‘**disallowable instruments**’. For more information about legislative instruments, see here.

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

### GLOSSARY IS OUT OF DATE

For the new key concepts see **APPENDIX: CHANGES IN KEY PROVISIONS — EXPOSURE DRAFT COMPARED WITH THE INTRODUCED BILL** at the end of this Guide.

# PART 2: DIGITAL DUTY OF CARE – PROS, CONS & AMBIGUITY (ISSUES FOR YOUR SUBMISSION)

This part covers the key issues in the Bill 2026. For each issue, we provide:

1. **an explanation of what the Bill does**
2. **pros and cons of the approach**
3. **areas of ambiguity and vagueness in the Bill.**

## 2.1 Single overarching duty – better than multiple duties?

The Bill imposes a single overarching duty: persons responsible for online services must ensure, so far as is reasonably practicable, a safe online environment (s 26(1)).

This is unlike the United Kingdom's Online Safety Act, which imposes multiple specific duties for specific types of harm.

The Australian approach draws on the model used in workplace health and safety law — a broad, principles-based obligation rather than a prescriptive checklist.

| Pros | Cons |
|---|---|
| Future-proofs the law - captures new technologies and harms without needing legislative amendments each time a new risk emerges. | May lack specificity - platforms may struggle to understand what compliance looks like in practice without detailed guidance. |
| Simpler for platforms to understand - one overarching standard to meet, rather than navigating a complex web of specific duties. | Could be interpreted narrowly by platforms seeking to minimise their obligations, undermining the protective intent of the law. |
| Avoids gaps that can emerge when regulation relies on a closed list of specific duties that may not anticipate every scenario. | The "reasonably practicable" qualifier to the duty to create a safe online environment may dilute the duty in practice - platforms could argue that addressing certain harms is not reasonably practicable due to cost or technical difficulty. |
| Consistent with other Australian safety frameworks (e.g. workplace health and safety), giving regulators and courts a well-established body of interpretive principles to draw on. | Without specific duties it may be harder for the regulator to take enforcement action and demonstrate non-compliance. |
| | The UK experience suggests that specific duties for specific harms provide clear accountability and more predictable compliance expectations. |

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

Platforms no longer have to ensure safety only "so far as is reasonably practicable", and the old test for what was reasonably practicable has gone, including its references to cost and privacy. Instead, platforms must "take appropriate action to ensure a safe online environment, including by taking effective measures to mitigate risks identified in risk assessments conducted in accordance with section 27" (Bill s 26(1)).

Whether an action is appropriate depends on four main factors (Bill s 26A):

- how far it actually makes the service safe;
- how serious the harmful material is, and how much the action reduces the risk;
- the impact on the security of the service; and
- whether the action is proportionate to the risk, judged by how likely the harm is and how serious it would be.

Crucially, the cost of an action must not be taken into account (Bill s 26A(3)).

### Rationale

The EM says the new test makes clear what platforms must do to meet the duty and allows penalties to be imposed. It also says a single piece of harmful content will not necessarily be a breach, but may be evidence of a wider, systemic problem.

Cost is excluded because "there are serious harms which platforms must take effective measures to address", and on a platform with millions of users, even a small percentage of harm can affect many people (EM p 67). This largely answers the concern above that platforms could use cost as an excuse not to act.

### Potential new advocacy points to consider

1. Support the rule that cost cannot be taken into account, and ask Parliament to keep it (Bill s 26A(3)).
2. Privacy is no longer one of the factors. It has been replaced by "the impact on the security of the online service" (Bill s 26A(1)(d)). The EM says the duty is not meant to undermine people's privacy, "including by disabling or dismantling features such as end-to-end encryption" (EM p 53), but the Bill doesn't say this. Privacy and digital rights organisations may want to ask for it to be written into the Bill.
3. "Appropriate" is still not defined beyond these factors, so the tip on unclear terms in section 2.3 still applies.

## 2.2 Protections for adults vs children – is a tiered approach the best approach?

The Bill creates a tiered protection structure with different levels of obligations depending on the age of the user:

- **Adults:** protected from "seriously harmful material and conduct" only (s 25C), with potential requirements for online services to provide adults with 'user empowerment tools' if required by the Minister (noting these requirements don't currently exist.)
- **Children:** Protected from material "harmful to children" (s 25D) and from harms arising from design features.
- **Under-16s on social media:** prohibits design features with "negative behavioural impacts".
- **Private communications exception:** the duty 'does not require any action to be taken in relation to lawful communication occurring in private solely between consenting adults' (the private communications exception) (s 26(7)).

| Pros | Cons |
|---|---|
| Children have distinct vulnerabilities and the well-established "best interests of the child" principle justifies heightened protections. | Creates an "age cliff" at 18 - a 17-year-old receives full protection, while an 18-year-old loses much of it overnight. |
| Reflects the greater capacity of adults to manage online risks and make informed choices about their online activity. | Many serious harms disproportionately affect adults: technology-facilitated gender-based violence, addiction targeting, elder abuse, and exploitation of people with disabilities. |
| The under-16 social media design feature prohibition responds to growing evidence of harm to young people from addictive design patterns. | The "seriously harmful", which requires material or conduct to be seriously harmful rather than just harmful, may be too high for many adults. |
| Consistent with international trends prioritising child safety, including the UK Online Safety Act. | Adults who are vulnerable to harms online (due to disability, age, addiction, or mental health conditions) may not be sufficiently protected by the digital duty of care. |
| The private communications exception seems to intend to protect adults from lawful and consensual conversation, implicitly recognising the rights to freedom of speech, expression, and privacy. | The private communications exception could have unintended consequences due to its wording (see below). It may also shield platforms from accountability for harms occurring in messaging services (e.g. coercive control, image-based abuse, sextortion). |
| | The UK removed its "lawful but harmful to adults" duty under political pressure, but many stakeholders including civil society organisations believe it should have been retained. |

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

The three tiers of protection remain (Bill s 25C), but what each tier covers has changed.

- **For everyone, including adults** (seriously harmful material, Bill s 25D):
    - The list is now closed, so the Minister can no longer add to it.
    - Four types of material have been added: fake nude images, sharing or threatening to share intimate images, material promoting gender-based violence, and violent extremist material.
    - Harassment is now caught if it is "reasonably likely" to cause serious harm, even if the harasser didn't intend that.
- **For children** (material harmful to children, now Bill s 25E):
    - This list is also closed, so the Minister can't add to it.
    - The catch-all now covers anything that "may cause serious harm to children".
    - Platforms that children could use to access pornography must check users' ages and restrict access (Bill ss 27D–27F).
    - The Act's objects now refer to "acting in the best interests of" children, not just "protecting" them (Bill s 3(b)).

### The private communications exception

This exception (now Bill s 26B) uses the same words as before, but the EM explains its limits:

- large group chats or servers, which may have hundreds or thousands of members, are not "private";
- unlawful communications are not protected; and
- tools such as child sexual abuse material detection, or tools that disrupt video and calls, "should be applied where there is a risk of serious illegal harm" (EM p 68).

### Rationale

The EM says the adult list is closed and closely matches material that is already illegal, to avoid "any disproportionate impact on material and conduct generally" (EM pp 54–57).

The children's catch-all relies on the Act's existing definition of "serious harm", which does not include "mere ordinary emotional reactions" such as distress or anger (EM p 59).

The age checks reflect requirements in the industry codes being repealed (EM p 76). The EM also argues that treating adults and children differently is justified (EM pp 20–21).

### Potential new advocacy points to consider

1. Harassment no longer has to be intended to cause serious harm. It is enough that it is "reasonably likely" to do so. Victims don't need to prove what the harasser meant, which may help adults experiencing technology-facilitated abuse. Consider supporting this change.

2. Adults are now protected from material promoting gender-based violence. Test the definition (Bill s 5) against your clients' experiences.
3. The age cliff at 18 remains.
4. On age checks (Bill s 27D), consider:
    - the privacy risks;
    - whether requiring platforms to offer just two ways to prove your age is enough. The Bill doesn't say what those methods must be, so both could require ID, or both could collect sensitive data such as a face scan (Bill s 27D(2)(a)); and
    - the impact on people without standard ID, LGBTIQ+ communities and sex workers.

    The Bill has no privacy safeguards for age checks. The EM says only that platforms are "not expected" to collect more personal information than they already do (EM p 18).
5. The EM says detection tools should be used in private messages, which raises concerns about scanning. Consider asking for the limits of the exception to be set out in the Bill.
6. Because the adult list is now closed, protections for adults can't be extended without changing the law.
7. Age checks are required only "where technically feasible and reasonably practicable" (Bill s 27D(2)). This is the same kind of qualifier that has been removed from the main duty. Consider asking whether it should be removed here too.

## 2.3 Design features – does the Bill go far enough?

The Bill addresses the role of platform design in causing or contributing to online harms:

- **Design features** are defined in s 25F and include: recommender systems (algorithmic recommendations), endless-feed (infinite scrolling), feedback mechanisms (likes, reactions), logged-in features (features requiring a login), and time-limited features (disappearing content). **Note:** Children are expressly protected from harms associated with design features (s 25B(1)(b)(ii), whereas for adults there are fewer protections.
- **Under-16s on social media:** design features with "negative behavioural impacts" (s 25G) must not operate at all, noting this only applies to social media.
- **User empowerment tools:** the Minister may require services to provide tools that allow users to control their experience (s 26(4)–(6)). This only applies where required by a Minister ie. through a legislative instrument.

- **Risk assessments:** must identify design features giving rise to foreseeable risks (s 26A).

| Pros | Cons |
|---|---|
| Directly addresses addictive design and algorithmic amplification, especially for children. This goes beyond content-only regulation to tackle the structural drivers of harm. | User empowerment tools that may be required to be made available to adults (if the Minister passes a legislative instrument to this effect) place the burden on individual users rather than requiring platforms to adopt safe defaults - this is a "safety opt-in" rather than "safety by design" approach. |
| The under-16 prohibition is a strong response to mounting evidence of design-driven harm to children and adolescents. | There is no requirement for recommender algorithms to be opt-in by default - users must actively seek out alternatives to algorithmic curation. |
| Risk assessment requirements force platforms to proactively consider the impacts of their design choices before harm occurs. | Critical details about user empowerment tools are left to delegated legislative instruments that have not yet been drafted, making it difficult to assess whether protections will be adequate. |
| User empowerment tools, if they are required by the Minister in future by a legislative instrument, would give individuals some degree of control over their online experience. | Does not address advertising and targeting design features - how data is collected and used to target advertisements is not covered by the design feature provisions. |
| | The listed design features may not capture all features causing harm (e.g. autoplay, push notifications, social comparison metrics, gamification mechanics). |
| | The under-16 prohibition is limited to social media services - other types of services (e.g. gaming platforms, streaming services) also deploy harmful design features. |

**Areas of Vagueness Requiring Clarification**

The design features provisions contain several significant ambiguities that organisations may wish to raise in submissions:

**User empowerment tools and ambiguity around opt-in or opt-out**

- Despite recent media coverage regarding opt-in or opt-out algorithms, the Digital Duty of Care doesn't require online services to provide this functionality to users.
- What it does require is that online services provide 'user empowerment tools' but only as required by the Minister (see s 26(3)(a) and (4)).

- ‘User empowerment tools’ allow users of online services to manage the way the design features of the service operate for them (for example, by providing control over an algorithm (s 26(6)).
- No ‘user empowerment tools’ are currently required to be provided under the Bill. What that means is, as the Bill stands, there would be no obligation for the service provider to give adults any control over the design features of online services, unless and until this was required by the Minister via a legislative instrument passed by Parliament.
- Organisations should consider whether online service providers should be required under the Bill to provide any ‘user empowerment tools’ and, if so, what tools (e.g opt-in or opt-out algorithms).

**Should additional design features be deemed as having ‘negative behavioural impacts’?**

- Section 25F lists five categories of design features (recommender, endless-feed, feedback, logged-in, time-limited).
- The use of the word “includes” in s 25F(1) demonstrates the list is non-exhaustive, meaning the duty could apply to design features not specifically listed, such as autoplay, push notifications, gamification mechanics, or social comparison metrics.
- In addition, certain design features of online services are deemed as having ‘negative behavioural impacts’ (that is, recommender features, logged-in features, endless feeds, feedback feeds, time limited features and any other design features determined by the Minister (see s 25G).
- Organisations should consider whether there are other design features that should be deemed to have ‘negative behavioural impacts’ other than those listed in s 25G(1), for example, autoplay, push notifications, gamification mechanics, or social comparison metrics. Otherwise, it will be up to the Minister and Parliament to pass a legislative instrument in order to identify other design features that are deemed to have ‘negative behavioural impacts’ which might take time or never happen.

**What constitutes “appropriate” management?**

- The mandatory requirement to “manage design features of the service appropriately” (s 26(3)(a)) does not define what “appropriately” means. This leaves significant uncertainty for both services (who must comply) and the Commissioner (who must enforce).
- Organisations may wish to recommend that the explanatory memorandum provide guidance on what “appropriate” management looks like in practice, or that the Commissioner be expressly empowered to issue binding standards on this point.

**TIP FOR YOUR SUBMISSION** Where the Bill is vague, recommend that the Government clarify its intent either in the explanatory memorandum, through amendments to the Bill itself, or by requiring the Commissioner to issue binding guidance.

## WHAT HAS CHANGED IN THE INTRODUCED BILL

1. **Which features count as harmful for children under sixteen.** The Bill no longer has a general definition of "design feature". For under-16s on social media, only three types of feature are now automatically treated as harmful (Bill s 25G(1)):
    - recommender features (algorithms that choose what users see);
    - "logged-in features"; and
    - any feature the Minister adds.
2. **The exposure draft listed five.** Endless scrolling, likes and follower counts, and disappearing content are still defined, but they now count only as "logged-in features". That means they count only where a user can see them "only when the end-user is using the service with an account" (Bill s 25G(4)). Read receipts are expressly excluded (Bill s 25G(7)).
3. **How the protection is worded.** Under-16s on social media must not "have access to" or be "exposed to" harmful features, rather than those features simply not operating (Bill s 25C(1)(c)). For children of all ages, the general protection now covers harms from "access or exposure to design features of online services" (Bill s 25C(1)(b)(ii)).
4. **Managing design features.** The requirement to "manage design features of the service appropriately" has gone from the duty. Risk assessments are now a separate obligation with their own penalty (Bill ss 27, 27B).
5. **User empowerment tools.** Tools that let users control their own experience are now in Bill s 27A. Before requiring one, the Minister must:
    5.1. be satisfied it is "necessary to protect users" from harms caused by design features (s 27A(3));
    5.2. not take the content on the service into account (s 27A(4)); and
    5.3. first seek the Commissioner's advice (s 28B(2)(d)).
6. **Looking inside the algorithms.** The Commissioner can now require platforms to explain how their algorithms make decisions and when a human reviews them. It can also require platforms to give an expert of its choosing access to examine the service (Bill s 28A).

### Does this mean children are only protected from logged-in features?

No. All children, on all online services, are still protected from harm caused by design features generally, whether or not they are logged in (Bill s 25C(1)(b)(ii)). The EM says platforms must consider how "algorithmic recommender systems, endless scroll and other potentially harmful or addictive features" could affect children (EM p 53). It also expects children to be protected when they are logged out (EM p 18), and says it does not matter whether users have an account (EM p 67).

What changes is the level of protection. For children under 16 on social media, endless scrolling, likes and follower counts, and disappearing content that can be used without logging in are no longer automatically banned. Instead, they fall under the general standard of "appropriate action" (Bill s 26A). The EM says platforms are not expected to remove potentially harmful features if the risk can be managed (EM pp 71–72).

## Rationale

The EM sets out evidence of harm from each type of feature: recommender algorithms, endless scrolling, likes and follower counts, and disappearing content (EM pp 62–66). It says endless scrolling includes "auto-play" (EM p 65).

However, it doesn't explain why endless scrolling, likes and disappearing content were taken off the list of features automatically treated as harmful for children under 16.

User empowerment tools are left to the Minister so the rules can keep up with technology, and the tools are meant to be "agnostic of the material served to users", meaning they control how features work rather than what content is shown (EM pp 74–75).

## Drafting inconsistency

Bill s 25G(4) defines a logged-in feature as the particular endless-feed, feedback or time-limited feature that is available only with an account. The EM (p 64) instead describes it at the level of the whole service: a service meets the condition if, for example, its feedback feature is available only when logged in, “regardless of whether those other features can be accessed while both logged-in and logged-out”. That wording reflects the service-level definition in ED s 25F(3). Organisations may wish to ask which reading is intended.

## Potential new advocacy points to consider

1. Ask for endless scrolling, likes and follower counts, and disappearing content to go back on the list of features automatically treated as harmful (Bill s 25G(1)). That way, children under 16 on social media would be protected from them whether or not they are logged in.
2. Ask for the EM's statements on endless scrolling, autoplay and protection when logged out to be written into the Bill.
3. Ask for the definition of "logged-in feature" to be clarified (see "Drafting inconsistency" above).
4. User empowerment tools are still up to the Minister, so the question of whether they will be opt-in or opt-out is still open. Recommending that recommender algorithms be switched off by default is still relevant.
5. The Bill no longer has a general definition of "design feature", so it is unclear whether features that aren't listed are covered. Consider asking for features such as autoplay, push notifications and gamification to be named in the Bill.
6. Consider supporting the Commissioner's new power to require platforms to explain how their algorithms work (Bill s 28A).
7. Children under 16 are not meant to have accounts on age-restricted social media platforms at all. So the logged-in rule will mostly matter where they have accounts anyway, or on services outside that regime. When they use a platform while logged out, only the softer general standard applies.

# 2.4 Scope of Harms — Are the Right Harms Covered?

The Bill defines two tiers of harm:

- **"Seriously harmful material and conduct" (s 25C)** categories include: material depicting serious violence, sexual exploitation, material promoting self-harm or suicide, material involving menacing, harassing or offensive conduct, material encouraging hostile attitudes towards women or gender equality, material promoting illicit drugs, and more. The Minister can add further categories by legislative instrument.
- **"Harmful to children" (s 25D)** categories include any material that could inflict serious harm to a child as well as certain categories of material and content harmful to children including: abuse, harassment or bullying, age-inappropriate content, content promoting eating disorders, and more, as well as any material determined by the Minister (via a legislative instrument).
- **Ministerial expansion power:** the Minister can expand both lists by disallowable legislative instrument, providing a mechanism for the framework to evolve as new harms emerge.

| Pros | Cons |
|---|---|
| Covers the most serious online harms based on current evidence, focusing regulatory attention where it is most needed. | Does not explicitly cover financial harms and scams (which are addressed under a separate regulatory framework), creating a gap for vulnerable people targeted by scam content on in-scope platforms. |
| Ministerial powers to add categories of harm allows the framework to evolve as new harms emerge without requiring primary legislation. | Does not address cumulative harms: for example repeated exposure to harmful content (e.g. body-shaming, radicalisation pathways) that may not meet the "seriously harmful" threshold individually but causes significant harm over time. |
| The Minister can only add categories of harm via legislative instruments which provide parliamentary oversight. Either House of Parliament can disallow a new category of harm or other changes the Minister wishes to make. | Does not explicitly address systemic harms: misinformation, disinformation, threats to democratic integrity, and coordinated inauthentic behaviour are not listed as harm categories. |
| The categories are technology-neutral - they apply regardless of how content is created or distributed (including AI-generated content). | Precursor harms (e.g. radicalisation pathwayss) may not be caught until harm has already occurred - the framework is reactive rather than preventive for these types of harm. |
| | The category "encourages or promotes hostile attitudes towards women or gender equality" may be difficult to enforce in practice - what constitutes |

| | |
|---|---|
| | "hostile attitudes" is inherently subjective and may be contested. |
| | Reliance on Ministerial expansion means new harms must wait for political action, which may be slow, subject to competing political priorities, and potentially impossible if parliament does not agree to the changes. |

**Areas of Vagueness Requiring Clarification**

**No clear threshold for "serious" harm.**

- The Bill distinguishes between "harm" (s 25E), "seriously harmful material and conduct" (s 25C) and material "harmful to children" (s 25D), but does not define where the threshold lies between harm and serious harm. This lack of a threshold creates uncertainty: when does distressing content cross the line from merely harmful to "seriously harmful"?
- For adults, only "seriously harmful" content triggers the specific protections in s 25B(1)(a). Without a clear threshold, services may under-moderate (arguing content does not meet the threshold) or over-moderate (removing content cautiously to avoid penalties).
- Organisations may wish to consider recommending that further guidance is provided (either in the explanatory memorandum, the Bill itself, or other guidance materials) on how seriousness is to be assessed, including whether cumulative or contextual factors are relevant.

**Risk of unintended consequences**

- Several of the harm categories in s 25C are broadly framed and could potentially capture conduct that is politically contested or subject to legitimate public debate. For example: s 25C(1)(j) covers material that "encourages, promotes, incites or provides instruction for the commission of a criminal offence" — this could potentially capture, for example:
    - content supporting or organising environmental protest actions (such as road blockades or trespass on industrial sites) that are technically criminal offences;
    - content discussing or supporting abortion access in jurisdictions where aspects of reproductive healthcare remain criminalised or contested; and
    - content expressing support for or organising protest activity that may be characterised as supporting a listed organisation.
- Similarly, s 25C(1)(k) covers material that "encourages, promotes, urges or instructs illicit drug use" — which could capture harm-reduction content (e.g. information about safe injecting practices) that is intended to reduce harm rather than promote drug use.
- Finally, it is unclear how the duty would apply to art, entertainment or satire depicting some of the harms.

- Organisations may wish to consider how such unintended consequences maybe relevant to their communities or client cohorts, and provide real life case studies in their submission to demonstrate their concerns.

  **TIP FOR YOUR SUBMISSION** Organisations should consider whether the Bill contains sufficient safeguards to prevent these provisions being applied to legitimate civic participation, protest, advocacy, or harm-reduction activities.

  Recommendations could include: requiring an intent element (that the material be intended to incite the criminal conduct, not merely discuss it); excluding material with a legitimate public interest purpose; or clarifying in the explanatory memorandum that these provisions are not intended to capture peaceful protest, political advocacy, or public health information.

## WHAT HAS CHANGED IN THE INTRODUCED BILL

The list of seriously harmful material is now in Bill s 25D, and the list of material harmful to children is now in Bill s 25E. Several changes respond directly to the concerns above.

- **Criminal offences.** The broad category covering any "criminal offence" now covers only serious federal offences, meaning those punishable by 3 years' imprisonment or more, that are "reasonably likely to cause serious harm" (Bill s 25D(1)(m)).
- **Illicit drugs.** This category has been removed.
- **No new categories.** The Minister can no longer add categories to either list.
- **Suicide.** This category now excludes voluntary assisted dying (Bill s 25D(1)(j)).
- **Excluded material.** A new category of "excluded material" falls outside both lists, so platforms do not have to act against it (Bill s 25F). It covers:
  - material connected with public officials' duties, and court and tribunal material;
  - professionally produced TV, radio, streaming, classified films and games;
  - professional news;
  - research;
  - literary, artistic or educational material that is not against the public interest;
  - public health, safety and harm-reduction information; and
  - anything "in the public interest".

  User-generated content inside computer games is not excluded (Bill s 25F(2)).
- **Defamation and financial loss.** The EM says the duty is not meant to cover defamatory content or content causing financial loss, unless it also meets one of the harm thresholds (EM p 53).

### Rationale

The EM says the narrower offence category targets "facilitation of offending, rather than the discussion of unlawful conduct" (EM p 56). It also says a closed list avoids overreach (EM p 57).

Excluded material is meant to protect "lawful and legitimate communications related to matters such as news reporting, professionally produced entertainment, or the public interest" (EM p 59).

On cumulative harm, the EM says risk assessments must consider campaigns of many small abusive messages from multiple users that may not seem serious on their own (EM p 72).

### Effect on this section

The examples of unintended consequences above (environmental protest, abortion access and harm reduction) are largely dealt with:

- most protest offences and State abortion laws are not serious federal offences;
- the illicit drugs category has gone; and
- excluded material protects harm-reduction, public health and public interest content.

The points above about the Minister adding to the lists no longer apply. However, scams, misinformation and other systemic harms are still not covered, and they now can't be added without changing the law.

### Potential new advocacy points to consider

1. Consider supporting the narrower offence category, the removal of the illicit drugs category and the new excluded material.
2. Test whether any serious federal offence relevant to your communities could still catch legitimate advocacy. Offences relating to terrorist organisations are one example.
3. "Professional news material" only covers outlets that have editorial independence and are bound by press council, broadcasting or similar standards (Bill s 25F(3)). Independent, community and First Nations media may have to rely on the broader "public interest" protection instead.
4. "In the public interest" is not defined. Consider asking for:
    - guidance and examples, such as peaceful protest, political advocacy and lived experience content; and
    - an explanation of how automated moderation will apply it.
5. Ask for the EM's statement on cumulative harm to be written into the Bill.
6. Consider whether a closed list is best, or whether there should be a way to add new harms with Parliament's approval.

## 2.5 Ministerial powers and legislative instruments

The Bill gives the Minister broad powers to make legislative instruments that shape and expand the regulatory framework quickly without needing to amend the primary legislation.

These instruments are disallowable, meaning they are subject to parliamentary scrutiny, and the Minister must seek the eSafety Commissioner's advice before making instruments under key provisions (s 27B) (but does not have to abide by the advice provided).

The Minister's instrument-making powers include the ability to:

- Exempt services from the definition of "online service" (s 25A(3))
- Expand the categories of "seriously harmful material and conduct" (s 25C(2))
- Expand the categories of material "harmful to children" (s 25D(2))
- Determine additional design features with "negative behavioural impacts" (s 25G(2))
- Determine what constitutes a "social media service" (s 25B(2))
- Require user empowerment tools (s 26(4))

| Pros | Cons |
|---|---|
| Provides necessary flexibility to respond to new technologies and emerging harms without the delay of amending primary legislation | The Bill uses a subjective "Minister is satisfied" test rather than objective criteria, giving the Minister wide discretion |
| Legislative instruments provide a parliamentary check on Ministerial power | There is no mandatory public consultation requirement before instruments are made — affected communities and industry may have no opportunity to provide input |
| The Commissioner consultation requirement (s 27B) helps ensure that expert, evidence-based input informs Ministerial decisions | Instruments take effect immediately, before the disallowance period expires, meaning a potentially problematic instrument can operate for weeks before Parliament has the opportunity to act |
| Consistent with other Australian regulatory frameworks (such as telecommunications and broadcasting) that routinely use delegated legislation to manage technical and operational detail | Disallowance motions are difficult to pass in practice in the Australian Parliament, as they require a vote in either House, reducing the practical effectiveness of this safeguard |
| | These powers could be used to narrow protections (through exemptions) as well as expand them, depending on the political priorities of the government of the day |

**Areas of Vagueness Requiring Clarification**

No clear process for removing or amending changes made by Minister via legislative instrument

- The Bill provides detailed powers for the Minister to make changes to how the Bill would operate in practice via legislative instrument (e.g. adding harm categories, design features, exemptions, etc.) but is silent on the process for removing or amending those changes once made.
- There is no requirement in the Bill for periodic review of changes made by the Minister via legislative instrument, no sunset clause (ie. automatic date it would expire), and no mechanism for affected parties to petition for the removal of a designation.
- This means that once the Minister adds a category of harm or a design feature to the list, it may remain indefinitely, even if circumstances change, evidence evolves, or the legislative instrument proves to have unintended consequences.
- Importantly, though, legislative instruments can generally be repealed/revoked by another instrument.

**Organisations may wish to make recommendations including:**

- mandatory periodic review of all changes passed via a legislative instrument (e.g. every 3–5 years)
- requiring the Minister to provide written reasons if they decide not to abide by advice provided by the eSafety Commissioner
- a formal petition mechanism allowing industry, civil society or affected communities to request removal or amendment of a change made via a legislative instrument
- sunset clauses requiring designations to lapse unless actively renewed
- public reporting by the Commissioner on the operation and impact of each legislative instrument.

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

The Minister's powers to add categories of harmful material, which were the main concern above, have gone. The Minister can still:

- exempt services that pose "little risk" (Bill s 25B(3)). The ground of being "used minimally in Australia" has been removed;
- decide whether a service counts as social media (Bill s 25C(2));
- add design features to the list of features treated as harmful to children under 16 (Bill s 25G(2));
- require user empowerment tools, but only if satisfied they are necessary (Bill s 27A); and
- make rules on researchers' access to platform data, if satisfied the access is reasonable (Bill s 205C).

The Minister must seek and consider the Commissioner's advice before using powers (a) to (d) (Bill s 28B). The Commissioner can also advise on emerging harms (Bill s 28C).

The Commissioner has also gained its own powers to set rules on:

- risk assessments (Bill s 27(4));
- complaint processes, after consulting the Privacy Commissioner (Bill s 27G); and
- what platforms must publish (Bill s 192L).

### Rationale

The EM says the exemption power lets the law focus on services that pose a real risk to Australian users (EM p 52).

It says requiring the Commissioner's advice "promotes due diligence" by making sure the Minister considers expert advice (EM p 80).

Leaving user empowerment tools to the Minister lets the rules keep pace with technology (EM p 74). Parliament can still reject these instruments (EM pp 53, 74).

### Potential new advocacy points to consider

1. The main concern above, that the Minister could expand the harm lists, has largely been resolved. Consider acknowledging this.
2. Exemptions can still narrow protections. For exemptions and the other remaining powers, consider recommending:
    - regular reviews;
    - sunset clauses, so that instruments lapse unless renewed; and
    - public consultation before they are made.
3. Recommend that the Minister give written reasons for not following the Commissioner's advice (Bill s 28B).
4. The Commissioner's own rule-making powers have grown. Consider recommending that civil society be consulted before they are used.
5. Because the harm lists are now closed, consider whether there should be a way to add new harms. One option is requiring Parliament to approve them before they take effect.

# 2.6 Rights protections

Notably, the Bill contains no express protections of human rights. This distinguishes it from comparable overseas legislation such as the EU Digital Services Act (which references fundamental rights) and the UK Online Safety Act (which expressly seeks to protect freedom of expression and privacy).

The Bill does include some features that indirectly address rights concerns:

- The "reasonably practicable" test (s 25H) balances cost and privacy impact, but does not expressly include freedom of expression considerations.
- The private communications exception for adults (s 26(7)) protects lawful private communications from the scope of the duty, implicitly protecting privacy rights and freedom of speech and expression.
- The harm categories in ss 25C and 25D define the scope of material that platforms must address, providing some boundaries on what is caught.

| Pros | Cons |
|---|---|
| The defined harm categories are narrow and mostly cover material that is already illegal or seriously harmful — the Bill does not create broad censorship powers (meaning any risks to freedom of speech or freedom of expression are ameliorated) | The absence of an express rights protection is a significant gap, particularly given the breadth of the duty and the severity of penalties |
| The systems-based approach (requiring risk assessments and safe design features) is less censorious than content-by-content removal, as it targets platform systems rather than individual speech | High penalties (60,000 penalty units) may incentivise platforms to over-remove content to avoid liability, creating a chilling effect on legitimate speech |
| Proportionality is built into the "reasonably practicable" standard — measures must be proportionate to cost and other factors | Some harm categories are inherently subjective (e.g. "hostile attitudes towards women or gender equality") and may be applied inconsistently across different platforms and contexts |
| The private communications exception seems to intend to protect private adult speech from regulatory reach (***although see analysis on ambiguity below***) | Automated content moderation systems produce false positives that disproportionately affect marginalised communities — including First Nations content, LGBTIQ+ content, and disability advocacy material |
| | There is no requirement in the Bill itself for platforms to provide reasons for content removal or to offer appeal mechanisms to affected users |

## Risk of unintended consequences

### *For civic participation and marginalised communities.*

The combination of broadly framed harm categories, high penalties, and no express rights safeguards (including freedom of speech or expression) creates a particular risk for communities whose speech is already subject to disproportionate suppression.

Aboriginal and Torres Strait Islander peoples, culturally and linguistically diverse communities, LGBTQI+ communities, disability advocates, and political activists may find that automated content moderation systems — deployed by risk-averse platforms seeking to avoid penalties — disproportionately flag or remove their content.

International research consistently shows that automated moderation tools produce higher false-positive rates for content in minority languages, content using reclaimed or in-group terminology, and content discussing sensitive topics such as racism, discrimination, or political protest.

Without an express freedom of expression safeguard and without requirements for human review of automated moderation decisions, the duty may inadvertently silence diverse and marginalised communities.

### *For private communications between adults*

The private communications exception provides that the digital duty of care 'does not require any action to be taken in relation to lawful communication occurring in private solely between consenting adults' (s 26(7)).

It is unclear how this would operate in practice, or whether online platforms would be expected to screen messaging services to ascertain or confirm that private communications are 'lawful', 'private', and 'solely between consenting adults'.

It is foreseeable that private communications between adults that are not lawful could be prevented from occurring. It is foreseeable that this might include, for example, conversations between adults relating to protest activity if the protest activity is criminalised in the state or territory in which they reside.

Organisations may wish to draw attention to the lack of rights protections in the bill by referring to 'real world' case studies, relevant to their clients and communities, that demonstrate the unintended consequences that might flow from a duty with no rights protections and broad categories of harm.

## WHAT HAS CHANGED IN THE INTRODUCED BILL

The Bill still has no express protection for freedom of expression or human rights. However, the indirect safeguards have shifted:

- **Privacy** is no longer a factor in deciding what action platforms must take (see section 2.1).
- **Security and proportionality** are now factors instead (Bill s 26A).
- **Excluded material** protects news, art, public health and public interest content (Bill s 25F).
- **The harm lists** are closed and narrower, including the offence category (see section 2.4).
- **Transparency.** The Commissioner can require platforms to publish information about content moderation, including automated moderation, what is removed and how many moderators they employ (Bill s 192L).
- **Complaints.** Complaint processes must be open equally to all individuals in Australia (Bill s 27G).
- The EM's human rights statement deals with freedom of expression, privacy, non-discrimination and children's rights (EM pp 9–22). It argues that the Bill targets platforms' systems and processes "rather than to restrict the sharing of individual pieces of content" (EM p 17).

### Potential new advocacy points to consider

- Consider acknowledging the new excluded material and the narrower harm lists.
- Continue to argue for an express protection of rights, given that breaching the duty carries a penalty of up to 60,000 penalty units (Bill s 26C).
- Test the EM's claim that the duty targets systems rather than individual pieces of content. Use case studies of automated moderation wrongly removing content.
- Recommend that the Commissioner's rules on complaint processes and on what platforms must publish (Bill ss 27G, 192L) require platforms to:
    - give reasons for moderation decisions;
    - offer appeals; and
    - publish data on automated moderation.
- Raise the EM's statement that detection tools should be used in private messages (EM p 68), and the fact that its promise on encryption is not in the Bill (EM p 53).
- Raise the privacy risks of age checks (Bill s 27D).

# 2.7 Enforcement and Redress

The Bill provides the eSafety Commissioner with a graduated enforcement toolkit but does not create a private right of action or an ombudsman or complaints body for individual online safety disputes (noting that the Commissioner still has capacity to receive complaints from users under the existing Online Safety Act infrastructure, which would remain in place even if the bill passes.)

The Commissioner's enforcement powers include:

- Formal warnings (s 26C)
- Remedial directions, with penalties of up to 60,000 penalty units for non-compliance (s 26D)
- Infringement notices
- Civil penalties of up to 60,000 penalty units for breach of the overarching duty (s 26B)

Other features of the enforcement framework:

- There is no private right of action — individuals cannot take direct legal action against platforms for breaching the duty
- There is no ombudsman or independent arbiter for individual disputes about online safety
- Platforms must maintain complaint and dispute resolution processes (s 26F)
- The Commissioner publishes annual monitoring and enforcement priorities (s 166C)
- The Commissioner can publish guidelines on the duty (s 26E)

| Pros | Cons |
|---|---|
| A centralised expert regulator ensures consistent, specialist decision-making — the Commissioner has deep expertise in online safety | Individuals harmed by platform failures cannot take direct action — they are entirely dependent on the Commissioner's discretion to act |
| The absence of a private right of action avoids vexatious or strategic litigation against online providers | There is no ombudsman for user disputes — if a platform's internal complaint process fails, there is no independent arbiter to turn to |
| The Commissioner can prioritise systemic issues over individual complaints, addressing root causes rather than symptoms | Commissioner resource constraints are a real concern — a single regulator may not have the capacity to address the volume of online harms experienced across Australia |
| The graduated enforcement toolkit (warnings, directions, penalties) provides flexibility to match the response to the seriousness of the breach | Platform complaint processes may be inadequate — many community organisations report that existing platform complaints mechanisms are slow, opaque, and ineffective |

| | |
|---|---|
| The annual priorities statement provides transparency about the Commissioner's enforcement approach | There is no requirement for the Commissioner to act on individual complaints, meaning some harms may go unaddressed |
| | The absence of a private right of action means the entire framework relies on public enforcement, with no safety net if the regulator is under-resourced or slow to act |
| | Some experts have recommended a multi-Commissioner model to share the regulatory burden, but the Bill retains the single-Commissioner approach |

## WHAT HAS CHANGED IN THE INTRODUCED BILL

1. **Penalties**
   - 1.1. The maximum penalty for breaching the duty stays at 60,000 penalty units (Bill s 26C).
   - 1.2. The penalty for ignoring a remedial direction, which is an order from the Commissioner to fix a problem, has dropped tenfold, from 60,000 to 6,000 penalty units (Bill s 26E; EM p 70).
2. **Infringement notices**
   - 2.1. These are fines a platform can pay instead of facing further enforcement action, such as court proceedings. They can now be used for all the new obligations, including remedial directions, but not for a breach of the duty itself (EM pp 81–82).
   - 2.2. For these obligations, the fine for companies rises from 60 to 600 penalty units (Bill s 163AA; EM pp 82–83).
3. **Separate enforcement for the new obligations**
   - 3.1. Each new obligation has its own formal warnings and remedial directions:
     - 3.1.1. risk assessments and user empowerment tools (up to 60,000 penalty units);
     - 3.1.2. ii. age checks (up to 60,000 penalty units); and
     - 3.1.3. iii. complaint processes (up to 6,000 penalty units).
   - 3.2. Complaint processes must be available "equally to all individuals in Australia", and the Commissioner must consult the Privacy Commissioner before setting requirements for them (Bill s 27G).
4. **New powers for the Commissioner**
   - 4.1. It can publish any formal warning (Bill s 183AA).
   - 4.2. It can require platforms to explain how their algorithms work, or to give an expert access to examine them (Bill s 28A).

4.3. It can use its information-gathering powers to check compliance with the Act generally, except the social media minimum age rules (EM pp 92–93).

**5. Limits and gaps**

5.1. Remedial directions can't require action that has no effect on Australian users. State and Territory agencies are excluded (Bill s 25A).

5.2. The duty now falls on whoever exercises "the greatest degree of day-to-day control" over a service, rather than anyone in a position to exercise day-to-day control (Bill s 25B(2)).

5.3. There is still no right for individuals to sue and no ombudsman.

## Rationale

The EM says the 60,000 penalty unit maximum matches competition and consumer law and reflects the resources of multinational platforms (EM p 68).

The Commissioner is expected to choose the right enforcement tool for the seriousness of the harm, and may sometimes go straight to court (EM p 69).

Opening complaint processes to people who don't use a platform means people such as deepfake or cyberbullying victims can seek a remedy even if they don't have an account (EM p 78).

Publishing warnings is about transparency (EM p 89), and platforms are not expected to hand over their algorithms (EM pp 79–80).

## Potential new advocacy points to consider

1. Ask why the penalty for ignoring a remedial direction was cut tenfold.
2. The EM's general description says infringement notices can be used to enforce the duty (EM p 69), but its detailed notes and the Bill don't allow this (EM pp 81–82). Consider asking for clarification.
3. Consider supporting complaint access for people who don't use a platform. Recommend that the Commissioner's rules for complaint processes (Bill s 27G) set minimum standards on:
    3.1. how quickly complaints are dealt with;
    3.2. giving reasons;
    3.3. review of decisions; and
    3.4. accessibility for culturally and linguistically diverse communities and people with disability.
4. Consider supporting the publication of formal warnings.
5. Continue to argue for an ombudsman or a right for individuals to sue.
6. The narrower definition of who is responsible may make senior staff less accountable. Consider raising this.

# 2.8 Transparency and Research

The Bill introduces a new transparency reporting regime and a data access scheme for approved researchers, together with novel protections for investigative research into platform behaviour.

The key elements are:

- **Transparency reporting (Part 12A):** The Commissioner can require platforms to produce transparency reports (s 192C) and can compel publication of safety information (s 192M)
- **Data access for researchers (Part 14):** An approved researcher scheme (s 205B) supported by data access rules (s 205C)
- **"Sock puppet" identity protections:** Researchers and the Commissioner are protected when creating test accounts to investigate platform behaviour (ss 205G–205L)
- Approved researchers must be employees of Australian universities

| Pros | Cons |
|---|---|
| The 'sock puppet' provisions are world-leading — they enable critical investigative research that would otherwise expose researchers to legal risk | Approved researchers are limited to employees of Australian universities — this excludes civil society researchers, independent researchers, journalists, and international academics who do important work on online harms |
| The data access scheme enables evidence-based policy by giving researchers access to platform data that is otherwise closely held | The data access rules are left to legislative instruments that have not yet been drafted — the practical detail of the scheme is unknown |
| The transparency reporting regime positions Australia alongside the EU in promoting platform accountability through public disclosure | There are genuine privacy and security tensions in providing researchers with access to platform data, and the Bill does not yet detail the safeguards |
| Transparency reports help the public, civil society, and policymakers understand platform safety performance and hold platforms to account | There is no requirement for platforms to provide data in interoperable or standardised formats, which may limit the usefulness of data access in practice |
| | Transparency reports may not be sufficiently detailed or comparable across platforms without minimum content standards |

**WHAT HAS CHANGED IN THE INTRODUCED BILL**

1. **Transparency reports** (now Bill ss 192B–192K). Some safeguards have been removed:
    - published reports no longer have to leave out confidential or personal information;

- the protection for confidential information that platforms are required to publish has gone; and
- the Commissioner has fewer matters to weigh up before requiring a report (Bill s 192B(3)).

2. **Access to platform data**. This is still limited to researchers employed by Australian universities, and the research must relate to online safety in Australia (Bill s 205B). The rules may give researchers access to information about algorithms and human review, but only to safety-relevant data. Researchers can't use the data for commercial gain (Bill s 205C).
3. **Sock puppet accounts.** These are anonymous test accounts that researchers and the Commissioner use to investigate how platforms behave.
    - Narrower protection. In the exposure draft, the protection overrode any other law. Now it overrides only contracts, such as a platform's terms of service (Bill ss 205H, 205J).
    - New testing power. The Commissioner can now test whether a service can be used to create illegal material, which the exposure draft didn't allow (Bill s 205K).
    - **Safeguards.** The Commissioner must set policies in consultation with the Australian Federal Police and publish an annual statement on its sock puppet activities (Bill ss 205K, 205M).

### Rationale

The EM says transparency reporting carries over the existing reporting obligations and extends them to the duty (EM pp 89–90).

Limiting data access to universities keeps sensitive data with qualified researchers working under ethics and governance standards (EM pp 95–96).

Sock puppet accounts are needed to test recommender algorithms and AI tools (EM pp 98–99).

The annual statement provides transparency without revealing the Commissioner's methods (EM pp 99–100).

### Potential new advocacy points to consider

1. The university-only rule still excludes civil society and independent researchers. Consider recommending an accreditation pathway for them, with ethics oversight.
2. Privacy organisations may want the removed confidentiality and privacy safeguards restored.
3. Because sock puppet protections no longer override other laws, researchers may still face legal risk. Consider asking for clarification.
4. Consider supporting the Commissioner's power to test whether services can create illegal material, given the safeguards of AFP-agreed policies and annual reporting.
5. Continue to recommend standardised and interoperable data formats.

# 2.9 New Take-Down and Link Deletion Powers

The Bill introduces several new powers aimed at the distribution layer of harmful content — targeting app stores, search engines, and the problem of reposted material.

**Part 6A — Fake nude material apps:** The Commissioner can require app stores and search engines to remove apps or websites that are designed for, or predominantly used to generate, fake nude material (s 86B). Platforms have a 24-hour compliance window, with penalties of 6,000 penalty units for non-compliance.

**Sections 72A and 92A— Link deletion for cyber-bullying material targeting children and adults:** The Commissioner can issue link deletion notices to search engines for cyber-bullying material targeting Australian children and adults. A removal notice must have already been issued as a prerequisite. Platforms have 24 hours to comply, with 6,000 penalty units for non-compliance.

**Complaint waiver:** The Commissioner can waive the requirement for a complaint to the provider where there is no functional complaints mechanism or where there is a risk of further harm to the victim.

**Reposting provisions:** The Commissioner can act on reposted material (even if altered) without requiring a fresh complaint, addressing the common problem of harmful material resurfacing in modified form.

| Pros | Cons |
|---|---|
| Addresses the distribution layer — links in search results can continue to surface harmful material even after the source content has been removed | Part 6A is limited to apps and websites whose "predominant purpose or use" is to generate fake nude material — many AI tools that can generate such material have broader purposes and would not be caught by this provision |
| The 24-hour compliance window ensures faster response times for victims | Link deletion is limited to search engines only — it does not cover social media sharing, messaging apps, or other distribution channels where harmful material circulates |
| The complaint waiver removes barriers for victims who cannot safely complain to the platform directly — for example, where doing so might alert a perpetrator | The sequential prerequisite for link deletion (a removal notice must be issued first) creates delay at a time when speed matters most |
| The reposting provisions address the persistent "whack-a-mole" problem, where harmful material is re-uploaded in slightly altered form | There is no individual complaint mechanism for Part 6A — only the Commissioner can initiate action |
| Part 6A directly targets the growing problem of non-consensual intimate image generation using AI tools | |

## WHAT HAS CHANGED IN THE INTRODUCED BILL

**Nudify apps.** The power to remove apps and websites that create fake nude images is narrower. Under the exposure draft, an app or website could be removed if it was either designed for this purpose or mainly used for it. The Bill requires both, and the app or website must also be available to users in Australia (Bill s 86B(1)). However, a removal notice can now also require app stores and search engines to take reasonable steps to remove copycat apps that are "substantively the same" (Bill s 86B(2)(d)).

**Reposted material.** The Commissioner can now act on reposted material without a fresh complaint at any time. The exposure draft's 6-month time limit has gone (Bill s 65(2C) and equivalent sections).

**Link deletion notices.** These now also cover intimate images (Bill s 82), and for the most serious (Class 1) material they can be issued at the same time as a removal notice (EM pp 33–35). A notice must still be "appropriate in the circumstances", and it must be withdrawn once it is no longer needed.

**Penalties.** The higher 6,000 penalty unit maximum for ignoring removal notices is kept. The exposure draft's increase in penalties for breaching the social media minimum age, from 30,000 to 60,000 penalty units, has been dropped (ED Schedule 2, items 13–15).

### Rationale

The EM says the nudify power targets "the worst of the worst enablers" of image-based abuse "with minimal impact on legitimate online services". It adds that the Commissioner may look at how a service is marketed, what tools it offers and what steps it has taken (EM pp 25–26).

Confusingly, the EM still describes the test as "designed… or is being predominantly used", which doesn't match the Bill's "and" (EM p 25).

Reposted material doesn't need to reappear on the same platform or be posted by the same person (EM p 29).

For link deletion, the Commissioner may consider whether a notice would make the material much harder to find (EM pp 32–33).

The higher penalties implement a recommendation of the 2024 review of the Act (EM pp 40–45).

### Potential new advocacy points to consider

1. The "and" test makes the nudify power narrower than both the exposure draft and the EM's own description. Consider asking for "or" to be restored.
2. Consider supporting the removal of the 6-month limit on reposts and the new power to target copycat apps.
3. There is still no power to remove the tools used to create other harmful deepfakes, although fake nude images are now seriously harmful material under the duty (Bill s 25D(1)(d)).
4. Consider asking why the increase in social media minimum age penalties was dropped.

# PART 3: CASE STUDIES – HOW THE DIGITAL DUTY OF CARE WOULD APPLY

The following worked examples illustrate how the Bill, if it became law, would apply to real-world scenarios. Each compares the current legal position with the proposed framework and identifies remaining gaps.

> **TIP FOR YOUR SUBMISSION** Organisations may opt to use a similar 'case study' structure to present their own case studies in submissions and demonstrates gaps or strengths of the Bill.

## Case study 1: Algorithmic Targeting Based on Addiction

### Hypothetical facts

A community legal centre assists clients experiencing alcohol addiction whose social media feeds are dominated by algorithmically targeted alcohol advertising. Platform algorithms appear to use data inferred from clients' online activity (e.g. searching for addiction recovery information) to target them with discounted alcohol content, undermining treatment and triggering relapse.

### Summary

This scenario is not within the remit of the current Online Safety Act. The Bill, if it became law, would require online services to ensure that children aren't targeted with an algorithm featuring alcohol advertising (although a one-off alcohol advertisement might not be covered), but the digital duty of care probably wouldn't apply to this scenario for adults (although this is unclear.) More detail in the table below.

| Topic | Online Safety Act (current) | The Bill (if it became law) |
|---|---|---|
| **Content** – alcohol advertising | ✗<br>Content involving alcohol content not within the remit of the Act. | ✗ For adults<br>Content involving alcohol would not be captured under the digital duty of care, does not fall within a category of 'seriously harmful material and conduct' (s 25C)<br><br>? For children<br>Online services might be required to ensure children do not see alcohol related content online, although this is not clear cut.<br><br>Although online services have a duty to (as far as is reasonably practicable) provide a safe |

| | | |
|---|---|---|
| | | online environment, which requires them to protect children from 'content that is harmful to children' (s 25B), it is not clear if alcohol advertising in and of itself is 'harmful to children' (s 25D(f)(i). |
| **Design features** – recommender feature advertising alcohol | <br>Recommender system (ie. targeted alcohol advertising) not within the remit of the Act. |  <u>For adults</u><br><br>It is unclear if online platforms would be required to do anything about this under their digital duty of care.<br><br>The recommender system is a defined "design feature" (s 25F) and, as part of the duty of care, online services must 'manage design features appropriately' *(the meaning of this is unclear)*<br><br>It could be argued that it is appropriate to managing such a design feature (ie. make sure it isn't used to target people experiencing addiction), but this is unclear.<br><br> <u>For children</u><br><br>Online services may be required to ensure (so far as is reasonably practicable) that children are protected from alcohol advertising via recommender systems under the digital duty of care.<br><br>This is because as part of their duty of care, online services must protect children from harms associated with design features, including recommender systems (s 25B(1)(b)(ii)). It is more likely that targeted alcohol advertising for children would be considered harmful to them. |

## Possible recommendations (ordered from most to least politically feasible)

- Expand harm categories to include content that exploits known vulnerabilities.
- Require recommender algorithms to be opt-in by default (aligns with the Government's announced policy direction on opt-out algorithms and has strong public support, but will likely face significant industry resistance).

- Include a specific prohibition on algorithmic targeting based on inferred health or addiction data (would require a provision and intersects with privacy reform - adds complexity to the Bill).
- Align with the EU Digital Services Act approach to advertising restrictions, including prohibiting targeted advertising based on special categories of personal data.

## Case study 2: The Age Cliff – How the Bill would apply to teens

### Hypothetical facts

Anja, who is 15-years-old, experiences sustained online harassment on a social media platform - including body-shaming, sexualised comments, and coordinated pile-on behaviour from anonymous accounts. The bullying she is experiencing is having significant mental health impacts.

She then turns 16, and the bullying and harassment continues.

Things get better for a couple of years but when she is in her final year of high school, and 18 years old, the co-ordinated online bullying starts again. Anja once again experiences significant mental health impacts and starts thinking about suicide.

### Summary

Under the Bill:

- At 15 years old, Anja benefits from the full suite of child protections under the Bill.
- At 16 years old, Anja gains access back to harmful features on her social media apps, but still receives some protection under the Bill.
- At 18 years old, the legal protections under the Bill change dramatically. Anja is not protected from the online abuse under the Bill.

| | Content | Design features |
|---|---|---|
| **15 years old** | Online services required to protect children from 'material and conduct that is harmful to children' including 'abuse, harassment or bullying' (s 25D(e))<br><br>Anja is protected under the Bill | Anja wouldn't have access to 'design features' on social media (such as algorithms that might be amplifying the abuse) because these are banned for under-16s (s 25B(1)(c)).<br><br>Anja is protected under the Bill |
| **16 years old** | As above<br><br>Anja is protected under the Bill | Because Anja has turned 16, she is now allowed 'design features' on social media, including algorithms.<br><br>But she is still protected somewhat because online services must protect |

| | | |
|---|---|---|
| | | children from 'harms associated with the operation of design features'<br><br>Anja is somewhat protected under the Bill |
| **18 years old** | Anja is now an adult. Online services don't have to protect adults from abuse, harassment or bullying because these are not considered to be 'seriously harmful content' (see s 25B(1)(a) and 25C.<br><br>Would only be protected if the bullying meets the higher threshold of intending to cause serious harm by menacing or harassing (s 25C(1)(f))<br><br>Anja is not protected under the Bill | Because Anja is now an adult, there are no specific protections relating to design features online including on social media.<br><br>Anja is not protected under the Bill |

**Possible Recommendations (ordered from most to least politically feasible)**

- Clarify in the explanatory memorandum that cumulative harassment - even if individual instances fall below the "seriously harmful" threshold - can constitute seriously harmful material or conduct when assessed as a pattern.
- Require risk assessments under s 26A to specifically identify risks to young adults aged 18 -25 as a distinct cohort, recognising their developmental proximity to children (minor addition to risk assessment requirements)
- Extend the design feature restrictions (or a graduated version of them) to young adults aged 18–25 on social media, creating a transitional period rather than a hard cliff (requires a new provision and a policy decision about the appropriate age boundary — will face arguments about arbitrariness and adult autonomy)

# PART 4: TEMPLATE SUBMISSION

Below is a template submission that your organisation can adapt. Text in [square brackets] should be replaced with your organisation's details. You may delete sections that are not relevant to your submission.

> **Formatting Tip** Government prefers submissions that are clearly structured, use headings, and include a summary of key points at the beginning (especially as the Secretariat may be using AI to assist with synthesising key themes from submissions). Keep your submission concise - quality over quantity.

*[Date]*

Digital Duty of Care Consultation
Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts

Via email: digitaldoc@infrastructure.gov.au

Dear Secretariat,

**Re: Submission on the Exposure Draft - Online Safety Amendment (Digital Duty of Care) Bill 2026**

*[Organisation name]* welcomes the opportunity to make this submission in response to the exposure draft of the *Online Safety Amendment (Digital Duty of Care) Bill 2026*.

## About [Organisation Name]

*[Organisation name]* is a *[type of organisation — e.g. community legal centre / not-for-profit / peak body / advocacy organisation]* based in *[location]*. We *[brief description of what the organisation does — e.g. provide free legal assistance to / advocate on behalf of / support] [description of communities served]. [Optional: We have been operating since [year] and assist approximately [number] people per year.]*

*[Organisation name]* has direct experience working with *[communities/individuals]* who are affected by online harms including *[brief description of relevant harms — e.g. technology-facilitated abuse, cyberbullying, algorithmic targeting, image-based abuse, online scams]*.

This submission draws on our frontline experience to *[support / raise concerns about / suggest improvements to]* the proposed Digital Duty of Care framework.

## Contact Person

*[Name]*

*[Title]*

*[Phone]*

*[Email]*

## Executive summary and recommendations

*[Insert a numbered list of your specific recommendations to Government, starting with a clear statement of whether your organisation supports the bill]*

1. *[E.g. Organisation name supports the introduction of an overarching duty as set out in section 26 of the exposure draft]*
2. *...*
3. *...*
4. *...*

## Detailed Submissions

*Flesh out your organisation's views selecting the issues where your organisation has relevant expertise and case studies to contribute. This resource details some of the issues your organisations could cover in detailed submissions, including:*

- *The Overarching Duty*
- *Protections for Adults vs Children*
- *Design Features*
- *Scope of Harms*
- *Ministerial Powers and Delegated Legislation*
- *Freedom of Expression*
- *Enforcement and Redress*
- *Transparency and Research*
- *New Take-Down and Link Deletion Powers*

*Use the structure from Part 3 to insert relevant case studies.*

*Set out any ideas your organisation has for future amendments to improve the bill or actions government can take to build on the duty and its enforcement, for example :*

- *a future statutory review of design features to add compulsion and addictive design, or*
- *a user empowerment tool that gives users the power to turn off behavioural advertising*

## Conclusion

*[Organisation name] [supports / broadly supports / has concerns about]* the introduction of a Digital Duty of Care. *[Brief concluding statement about the importance of the reform and what it means for the communities you serve.]*

We would welcome the opportunity to discuss this submission further.

Yours faithfully,

*[Name]*

*[Title]*

*[Organisation name]*

# APPENDIX: QUICK REFERENCE - KEY PROVISIONS

The following table summarises the key provisions discussed in this resource.

| Provision | What It Does | Section Reference |
|---|---|---|
| Overarching duty | Requires persons responsible for online services to ensure, so far as is reasonably practicable, a safe online environment | s 26(1) |
| Online service | Broad definition: social media, search engines, app stores, hosting, devices, AI content generation | s 25A |
| Safe online environment | Tiered protection: all persons (seriously harmful), children (harmful to children + design features), under-16s (design feature prohibition) | s 25B |
| Seriously harmful material and conduct | Defined categories of material/conduct harmful to all persons in Australia | s 25C |
| Harmful to children | Broader categories of material harmful to children | s 25D |
| Design features | Recommender, endless-feed, feedback, logged-in, time-limited | s 25F |
| Negative behavioural impacts | Design features that negatively affect children on social media | s 25G |
| Reasonably practicable | Balances harm likelihood/degree, knowledge, mitigation availability, cost, and privacy | s 25H |
| Risk assessments | Mandatory identification and assessment of foreseeable risks; annual review | s 26A |
| Breach of duty | Civil penalty: up to 60,000 penalty units | s 26B |
| Formal warnings | Commissioner may issue formal warning for breach | s 26C |
| Remedial directions | Commissioner may direct specific actions; 60,000 penalty units for non-compliance | s 26D |
| Commissioner guidelines | Commissioner may publish guidelines on the duty | s 26E |
| Complaint processes | Platforms must maintain complaint and dispute resolution processes | s 26F |

| | | |
|---|---|---|
| Private comms exception | Duty does not apply to lawful private communications between consenting adults | s 26(7) |
| Ministerial instruments | Minister must seek Commissioner's advice before making key instruments | s 27B |
| Fake nude material | Commissioner can require removal of apps/sites for generating fake nude images; 24-hour compliance | Part 6A, s 86B |
| Link deletion (children) | Link deletion notices for cyber-bullying material targeting children (search engines) | s 72A |
| Link deletion (adults) | Link deletion notices for cyber-abuse material targeting adults (search engines) | s 92A |
| Transparency reports | Commissioner can require transparency reports and compel publication of safety information | Part 12A |
| Data access | Approved researcher scheme with data access rules | Part 14, s 205B |
| Sock puppet protections | Identity protections for researchers and Commissioner using test accounts | ss 205G-205L |
| User empowerment tools | Minister may require services to provide user control tools | s 26(4)-(6) |
| Enforcement priorities | Commissioner must publish annual statement of monitoring and enforcement priorities | s 166C |
| Repeal of codes/standards | Repeals existing industry codes and standards (Schedule 3) | Schedule 3 |

# APPENDIX: CHANGES IN KEY PROVISIONS — EXPOSURE DRAFT COMPARED WITH THE INTRODUCED BILL

This Appendix compares the key provisions of the exposure draft (ED, 7 September 2026) with the Bill as introduced into the House of Representatives (Bill), with references to the Explanatory Memorandum (EM). The quick reference table that follows uses exposure draft section numbers.

| Provision | Exposure draft (ED) | Introduced Bill | EM reference |
|---|---|---|---|
| **Overarching duty** | s 26(1): ensure, so far as is reasonably practicable, a safe online environment | s 26(1): take appropriate action to ensure a safe online environment, including effective measures to mitigate risks identified in risk assessments | p 67 |
| **Reasonably practicable / appropriate and effective** | s 25H: likelihood, degree, knowledge, available ways, cost (grossly disproportionate) and privacy | s 26A: extent a safe online environment is achieved, nature of material, risk reduction, security of the service, proportionality to risk; cost must not be taken into account (s 26A(3)) | pp 67–68 |
| **Online service** | s 25A(1): any internet carriage service; equipment "in connection with" listed services; services allowing users to generate AI material and share it via another service | s 25B(1): internet carriage services supplied to the public; equipment enabling communications using an internet carriage service; any service that "generates material using artificial intelligence"; State and Territory agencies excluded (s 25A(1)) | pp 51–52 |
| **Person responsible** | s 25A(2): provider and a person "in a position to | s 25B(2): provider and, except for equipment services, the person | p 52 |

| | | | |
|---|---|---|---|
| | exercise day-to-day control” | exercising “the greatest degree of day-to-day control” | |
| **Exemptions** | s 25A(3): little risk or used minimally in Australia | s 25B(3): little risk only; Commissioner’s advice required (s 28B) | p 52 |
| **Safe online environment** | s 25B: protection from harms associated with “the operation of” design features; under-16 features “do not operate” | s 25C: protection from harms associated with “access or exposure to” design features; under-16s on social media “do not have access to, and are not exposed to” features with negative behavioural impacts | pp 52–53 |
| **Seriously harmful material** | s 25C: any “criminal offence” (j); illicit drug use (k); Minister may add categories ((m), (2)); menacing or harassing material “intended to” cause serious harm (f) | s 25D: exhaustive list; “serious Commonwealth offence” reasonably likely to cause serious harm (m); drug limb removed; adds fake nude material (d), intimate images (e), gender-based violence (h), violent extremist material (n); menacing or harassing material “reasonably likely to” cause serious harm (i); voluntary assisted dying carve-out (j) | pp 53–57 |
| **Harmful to children** | s 25D: catch-all for material that “could inflict serious harm” or that the Minister determines | s 25E: catch-all for material that “may cause serious harm to children”; no Ministerial power; “seriously harmful practices” | pp 57–59 |
| **Excluded material** | None | s 25F: official, court, professional entertainment, professional news, research, artistic and | pp 59–61 |

| | | | |
|---|---|---|---|
| | | educational, public health, harm reduction and public interest material (user-generated game content not excluded) | |
| **Design features and negative behavioural impacts** | s 25F: "design feature" includes five features; s 25G: all five, plus features the Minister determines, have negative behavioural impacts | No general definition; s 25G(1): recommender features, logged-in features (endless-feed, feedback or time-limited features accessible only with an account) and features the Minister determines; read receipts excluded (s 25G(7)) | pp 61–66 (logged-in feature p 64) |
| **Mandatory requirements** | s 26(3): manage design features appropriately, conduct risk assessments and take effective measures; failure is a breach of the duty | Risk assessments are a standalone obligation (s 27; penalty s 27B, 60,000 penalty units); "manage design features appropriately" no longer appears | pp 70–74 |
| **Risk assessments** | s 26A: "all reasonably foreseeable risks"; "content"; reassessment before "any changes" | s 27: "reasonably foreseeable risks"; "material or conduct"; reassessment before "significant changes"; 6-year retention; provided to the Commissioner within 30 days of a request | pp 70–74 (cumulative harms and AI p 72) |
| **User empowerment tools** | s 26(4)–(6): the Minister may require tools | s 27A: the Minister may require tools and prescribe requirements; must be satisfied a tool is necessary (s 27A(3)); must not take content into | pp 74–75 |

| | | | |
|---|---|---|---|
| | | account (s 27A(4)); Commissioner's advice required (s 28B) | |
| **Private communications exception** | s 26(7) | s 26B (same wording) | p 68: large group chats are not private; unlawful communications are not covered; detection tools should be applied where there is a risk of serious illegal harm |
| **Age assurance and access controls** | None | ss 27D–27F: where there is a foreseeable risk of children accessing pornography or fetish material, a choice of at least 2 age assurance measures plus access control measures (60,000 penalty units) | pp 76–77 |
| **Penalties and enforcement** | Duty breach s 26B (60,000 penalty units); remedial direction s 26D (60,000 penalty units); infringement notices cover breach of the duty (ss 163(1)(a), 163A) | Duty breach s 26C (60,000 penalty units); remedial direction s 26E (6,000 penalty units); infringement notices do not list s 26C (ss 163(1), 163AA); Commissioner may publish any formal warning (s 183AA) | pp 68–70, 82 |
| **Complaint processes** | s 26F: available equally to all Australians | s 27G: available equally to all individuals in Australia; Commissioner must consult the Privacy Commissioner | p 78 |
| **Algorithm information powers** | None | s 28A: notices requiring information about features, algorithms and human review, | pp 79–80 |

| | | | |
|---|---|---|---|
| | | and access for a nominated expert (6,000 penalty units) | |
| **Ministerial advice requirement** | s 27B: exemptions, both harm lists, design features | s 28B: exemptions, social media determinations, design features, user empowerment tools | p 80 |
| **Transparency** | ss 192C–192P, including restrictions on publishing confidential and personal information (s 192G(2)) and a confidentiality carve-out (s 192M(6)) | ss 192B–192N; those restrictions and the carve-out have been removed | pp 89–92 |
| **Data access and sock puppets** | ss 205B–205L: university researchers; sock puppet protection overrides any other law; Commissioner may generate only material that is not an offence | ss 205B–205M: university researchers; algorithm data; no commercial use; protection overrides only agreements, contracts and deeds; Commissioner may test whether illegal material can be generated; AFP-consulted policies and an annual statement | pp 95–100 |
| **Fake nude material apps** | s 86B(1): designed for the predominant purpose "or" predominantly used | s 86B(1): designed for the predominant purpose "and" predominantly used; "substantively the same" apps (s 86B(2)(d)) | pp 25–26 (describes the test as "or") |
| **Reposted material** | s 65(2C)(a): prior removal notice within the previous 6 months | s 65(2C)(a): any previous removal notice (also ss 66, 88, 89 and 90) | p 29 |
| **Points of contact** | s 229A: Australian-resident contact; determinations | s 229A: Commissioner may require a person responsible to | pp 101–102 |

| | | | |
|---|---|---|---|
| | by class; 1,000 penalty units | nominate a point of contact by notice | |
| **Commencement** | Schedule 1: day after Royal Assent; Schedules 2 and 3: 12 months after Royal Assent | Unchanged | p 2 |

# Authors




Leanne Ho
Partner
Leanne.Ho@wottonkearney.com



Claire Thurstans
Special Counsel
Claire Thurstans@wottonkearney.com


Wotton Kearney developed this resource in partnership with Digital Rights Watch.

14 September 2026