Inquiry into the Online Safety Amendment (Digital Duty of Care) Bill 2026

eSafety submission

November 2026

Contents

Introduction 2
Table 1: Enforcement pathways under the proposed digital duty of care. 4
Key positions and considerations 5
Appendix A: Strengthening the digital duty of care framework 13
Appendix B: Risk assessments 15
Appendix C: Nudify services 17
Appendix D: Implementation considerations 18
Appendix E: Testing accounts (sock puppets) 21
Appendix F: Approved researcher data access schemes 24

Content warning: This submission contains references to suicide, self-harm, child sexual exploitation and abuse, violence and other themes some people may find distressing. Please take care when reading.

Introduction

The eSafety Commissioner (**eSafety**) welcomes the opportunity to provide a submission to the Environment and Communications Legislation Committee.

eSafety is Australia's national, independent regulator and educator for online safety. Our remit, powers and functions are established under the *Online Safety Act 2021* (Cth) (**Online Safety Act**).

eSafety acknowledges the significance of the proposed reforms in the *Online Safety Amendment (Digital Duty of Care) Bill 2026* (**the Bill**) and the strong interest in the introduction of a digital duty of care.

eSafety engaged closely with Ms Delia Rickard PSM in her independent Statutory Review of the Online Safety Act (**Independent Review**),[1] which recommended that to enhance online safety for all Australian's, a singular and overarching duty of care be adopted that encompasses due diligence, underpinned by safety by design principles, risk assessment, mitigation and measurement. The Senate Environment and Communications References Committee has also previously recommended that the Australian Government legislate a digital duty of care to make online platforms safer for all users.[2]

We have also worked closely with the Department of Infrastructure, Transport, Regional Development, Communities, Sport and the Arts (DITRDCSA) to share our regulatory insights and expertise to inform the Bill. eSafety has had regular engagement with DITRDCSA including participating in legislative design workshops and providing feedback on draft iterations of the Bill. Through this process, eSafety has provided advice to DITRDCSA to identify areas which could be strengthened, focusing most critically on matters related to regulatory clarity and enforceability.

eSafety notes the Government's continued commitment to strengthening the Online Safety Act. eSafety welcomes the proposed enhancements to the complaints schemes, the introduction of increased penalties for non-compliance, the streamlining of compliance and investigations powers, and expanded ability to disclose information to Commonwealth entities.

A well-designed digital duty of care presents an important opportunity to strengthen Australia's online safety framework. eSafety supports the policy intent underpinning the

---

[1] Delia Rickard PSM (2024) *Independent statutory review of the Online Safety Act*, DITRDSCA website, accessed 5 November 2026.
[2] Senate Standing Committees on Environment and Communications (2025) *Internet Search Engine Services Online Safety Code and under 16 social media ban*, Recommendation 2, Parliament of Australia website, accessed 5 November 2026.

digital duty of care, including its focus on driving more systemic, proactive and preventative approaches to online safety, in line with our longstanding safety by design principles.

The Bill covers seriously harmful material and conduct, and material and conduct harmful to children, and targets harmful design features. This strengthens the existing scope of the Online Safety Act, which outside the social media minimum age provisions, is currently limited to material and content, generally not applicable to conduct or activity, and relies on definitions of harmful material determined with reference to the National Classification Scheme. eSafety is pleased to see that the harms captured under sections 25D and 25E of the Bill, as introduced to Parliament, were not tied to definitions of Class 1 and Class 2 material, which are defined with reference to the National Classification Scheme. These definitions have caused operational challenges and limitations for eSafety, particularly noting the National Classification Scheme was not designed with online material in mind, including user-generated and live-streamed content, but rather for professionally produced materials including films, books and cultural products. This change recognises that over the last decade, eSafety's investigators have developed significant expertise rapidly assessing user-generated and live-streamed content. Removing definitions of content and conduct linked to the National Classification Scheme is a significant improvement which, in addition to enhancing protections for Australians, will streamline and future-proof the Online Safety Act.

Given the scope of the Bill, this submission focuses on areas that would strengthen the legislation to ensure the proposed measures are the most effective, enforceable and likely to achieve the intended policy outcomes. This includes the proposed digital duty of care framework and related provisions concerning researcher access to data, the use of testing (sock puppet) accounts, transparency, and nudify services. This submission also outlines implementation considerations such as resourcing and capability readiness requirements, to ensure eSafety is sufficiently equipped to effectively enforce the Bill.

eSafety has developed significant regulatory experience through administering and enforcing novel regulatory schemes, including the Online Safety Codes and Standards, the Basic Online Safety Expectations, complaints schemes, and the Social Media Minimum Age, and championing the world-leading Safety by Design initiative. It is critical that the digital duty of care builds on the lessons learned through the implementation of the Online Safety Act to date.

eSafety's focus is on ensuring that the digital duty of care builds on the strengths of the existing framework, maintains minimum safety requirements, delivers practical and enforceable regulatory outcomes, and ultimately contributes to a safer online environment for all Australians.

eSafety is happy to provide any further information that would assist the Committee.

Table 1: Enforcement pathways under the proposed digital duty of care

To further contextualise eSafety's key recommendations and considerations below, it is important to understand that the Bill allows for four regulatory enforcement pathways under the digital duty of care framework.

Non-compliance with any of the below requirements is subject to 60,000 penalty units or $109.2 million for a body corporate. eSafety may also issue formal warnings which can be published, or remedial directions, for which non-compliance may result in 6,000 penalty units or up to $10.9 million for a body corporate.

Requirement	Enforcement pathway
Overarching digital duty of care (section 26)	eSafety must establish whether a service has taken appropriate action to ensure a safe online environment, including by taking effective measures to mitigate risks as identified in risk assessments. Determining whether a service has complied with this obligation includes considering a range of matters under section 26A, including the extent to which the action or measures achieve, or would achieve, a safe online environment, the impact on security and proportionality to risk. Effective mitigation of risks relates to risks which are identified in risk assessments conducted under section 27.
Risk assessments (section 27)	eSafety must establish whether a service has conducted a risk assessment and whether the risk assessment meets requirements under section 27 and any legislative instrument under section 27(4).
User empowerment tools (section 27A)	eSafety must establish whether a service has complied with requirements regarding user empowerment tools as prescribed under Ministerial legislative instrument.
Age assurance and access control measures (section 27D)	eSafety must establish whether a service has complied with requirements to implement appropriate age assurance and access control measures. This would include establishing that there was a reasonably foreseeable risk that children may access or be exposed to pornography or fetish material, and that the age assurance and access control measures were both technically feasible and reasonably practicable.

Key positions and considerations

This section outlines key positions and considerations that eSafety considers important for the Committee and stakeholders to note in relation to the proposed digital duty of care framework. These observations are intended to support a clear and consistent understanding of the framework, respond to matters raised through parliamentary debate, public discussion and stakeholder feedback, promote aligned expectations ahead of implementation and enforcement, and further assist eSafety in providing effective regulatory oversight while delivering greater clarity and regulatory certainty for industry. Where appropriate, further detail is contained in the appendices.

1. **Should Parliament wish to consider enforceable baseline obligations for industry, the Bill could enable minimum standards to be prescribed by way of a disallowable legislative instrument.**

eSafety acknowledges there has been Parliamentary debate and stakeholder concern regarding the repeal of the Online Safety Codes and Standards once the digital duty of care is in effect. As currently drafted, the Bill does not provide any mechanism to set specific obligations. While eSafety will provide regulatory guidance that outlines what it sees as best practice, this guidance is not enforceable.

Noting this debate, should Parliament seek to establish enforceable baseline obligations under the digital duty of care, this could be achieved by amending the Bill to allow for minimum obligations to be prescribed under a disallowable legislative instrument. Prescribing under legislative instrument is preferable to embedding obligations in the Act, as legislative instruments are more flexible to amend and therefore more future-proofed in line with the intent of the digital duty of care framework. Any legislative instrument would be subject to Parliamentary scrutiny and a disallowance period, providing stronger transparency and oversight by government than the current Codes which are developed by industry and registered by the Commissioner.

The combination of an overarching duty of care combined with specific obligations would establish both a proactive onus on industry to take responsibility for ensuring a safe online environment and a meaningful change in how industry approaches the design, development and deployment of their services into Australia, while also providing prescriptive and enforceable baseline obligations, especially for the most serious of harms. Online safety regulation in comparable international frameworks such as Ireland, Singapore, the United Kingdom, and the EU, commonly introduce regulator-developed codes or embed specific minimum obligations in legislation to complement risk assessment requirements.

For more detail, refer to **Appendix A: Strengthening the digital duty of care framework**.

2. **eSafety supports the inclusion of the risk assessment framework, which will be a key mechanism for operationalising the digital duty of care by providing a documented record against which eSafety can assess compliance with the duty. The risk assessment framework could be strengthened by clearly requiring the first risk assessment to be completed by commencement of the digital duty of care, and by providing eSafety with additional tools to address inadequate assessments and monitor compliance, including the ability to:**
    - **require re-assessment;**
    - **require an independent audit at the service's expense; and**
    - **give services an information notice which would require recipients to take steps to allow a person authorised by eSafety to remotely view certain information from the platform in real time, such as tests and demonstrations.**

eSafety supports the proposed risk assessment framework, recognising that it creates an important proactive obligation on industry to identify, assess and mitigate the risk of harms arising from their services. A robust risk assessment process has the potential to drive meaningful safety improvements by embedding risk management into service design, governance and decision-making.

eSafety notes the Bill provides the ability to issue remedial directions to services if they fail to conduct risk assessments in accordance with section 27. eSafety considers this could be strengthened by providing eSafety with the express ability to require re-assessment to prevent against malicious compliance, whereby a service may provide lengthy risk assessments without meaningful substance, minimising or mischaracterising risk ratings. The power to address inadequate risk assessments in a timely manner, where the risk assessment and proposed mitigation measures fail to address significant safety concerns identified by eSafety, would facilitate compliance with the risk assessment framework and reduce the risk of a burdensome and resource intensive administrative process that could potentially delay safety uplift.

The following measures would also strengthen eSafety's ability to verify information in risk assessments and monitor compliance:

- **The ability to require services to undertake an independent audit at their own expense** if the Commissioner has reason to believe the service is not complying with its obligations under the digital duty of care or its risk assessment is inadequate. Independent audits would provide an important mechanism for verifying providers'

claims, assessing the effectiveness of risk management measures and identifying systemic compliance issues. Given the scale and technical complexity of online services, external auditors provide expertise and resources where eSafety may not have direct access to the technical systems, data, or specialist capabilities needed to verify risk assessments and compliance, or a company may have concerns about sharing commercially sensitive information through the risk assessment process and need a method to verify their claims. In requiring a third-party audit, the Commissioner must have regard to the degree of risk posed by the service as a result of the likely harms and number of Australian users. eSafety notes that high-risk services can be of varying sizes. eSafety is committed to ensuring this does not impose unnecessary regulatory burden or expenses on regulated entities who are complying with their obligations but provides a critical tool to further validate platform claims and provide platforms an additional mechanism to demonstrate compliance.

- **The ability to give services an information notice which would require recipients to take steps to allow a person authorised by eSafety to remotely view certain information from the platform in real time, such as tests and demonstrations**. This amendment would ensure eSafety has sufficient powers to effectively verify provider's claims regarding the safety measures they have implemented and to assess the practical effectiveness of risk mitigation measures in operation by enabling direct observation of product functionality, experiences and risks. This is particularly important in an increasingly opaque environment and where services' claims and responses alone are insufficient. It would also enhance eSafety's investigative and enforcement capabilities in relation to emerging and rapidly evolving technologies by enabling direct observation of product functionality, user experiences, and associated online safety risks. eSafety notes this power is available to the United Kingdom's regulator, Ofcom, under section 100(3) of the *Online Safety Act 2023* (UK). Ofcom may issue a remote viewing information notice requiring a regulated service provider to allow Ofcom to remotely observe specified information, systems, processes or functionality relevant to Ofcom's online safety functions. The purpose of this notice is to help Ofcom understand how a service operates and assess compliance with statutory duties.

Initial implementation of the digital duty of care could also be strengthened if the legislation explicitly requires that the first risk assessment be completed by commencement of the digital duty of care. We note the digital duty of care and its requirements would not come into effect until 12 months from royal assent. If it is the governments' intent that services should complete risk assessments in anticipation of the duty coming into effect at 12 months, and that the eSafety Commissioner can request risk assessments upon

commencement of the duty to be provided within 30 days, this could be made clearer in the Bill. To avoid any prevarication or challenges by industry regarding the reasonableness of timeframes, it would be beneficial to require that services must have risk assessments in place by the time the duty commences. This would also give industry as much notice as possible and greater clarity around expectations upon commencement of the duty.

For more detail, refer to **Appendix B: Risk assessments.**

3. **Where new terms are not defined, such as harms under sections 25D and 25E of the Bill, eSafety recommends definitions be determined which can be debated and agreed to by Parliament to give the regulator the clearest approach to enforcement and to provide industry with sufficient clarity.**

eSafety notes the Bill covers seriously harmful material and conduct, and material and conduct harmful to children, and targets harmful design features, which strengthens the existing scope of the Online Safety Act. The Bill introduces new terms which are not expressly defined elsewhere, including harms such as material or conduct that encourages or promotes hostile attitudes towards women or gender equality (section 25E(1)(c)). eSafety notes this harm is a significant concern, with our research showing that children are frequently exposed to content and conduct that promotes hostility, prejudice and discrimination based on gender, and many children directly experience gender-based abuse and harassment online. In addition, children's exposure to violent pornographic content may contribute to the normalisation of sexual violence, coercion and other forms of gender-based violence against women and girls. eSafety's 2025 survey of 3,454 Australian children aged 10 to 17 found that:

- 47% had seen or heard offensive, sexist or hurtful things online about girls or women (37% in the last 12 months).
- 35% of children had seen online hate directed at women or girls (24% in the last 12 months).
- A significant minority of children (12%, 8% in the past 12 months) had been exposed to violent sexual images online.

Given the significance of this harm, eSafety recommends that new key terms be defined in legislation. Definitions settled by Parliament would provide greater certainty for industry to support consistent compliance with the digital duty of care.

4. Section 86B could be strengthened to empower eSafety to give notices to infrastructure services such as internet service providers, hosting service providers and pass-through services, requiring them to cease providing services or delivering material in relation to nudify platforms.

eSafety notes the Independent Review recommended that Government explore how best to prohibit search engines and app stores from surfacing, selling or distributing nudify services (recommendation 27).[3]

Based on eSafety's experience of how nudify services are accessed and used, including action taken against 8 of the most commonly accessed nudify services in Australia, restricting the availability of nudify services could be strengthened by expanding the scope of section 86B across the tech stack. With regards to nudify websites, as currently drafted, the removal notice powers would only require search engine services to remove links to the nudify website. While this may reduce the accessibility of the website, the website would still remain available to those who know its URL. Additionally, eSafety has observed that providers of nudify services often 'phoenix', including by duplicating substantially identical services under new domain names, URLs or other online locations. Providing eSafety with the ability to direct internet service providers, hosting service providers and pass-through services to cease supporting identified nudify services would enable more timely and durable disruption of harmful services, and ensure that responsibility for addressing these harms is more appropriately shared across the entities that facilitate their operation and availability.

For further detail refer to **Appendix C: Nudify services.**

5. The Bill represents an evolutionary shift in regulatory approach from the current Online Safety Act, which will necessitate a change in enforcement approach.

Whilst the introduction of an overarching digital duty of care provides a more flexible, durable and proactive framework for addressing online harms, this approach fundamentally changes the nature of regulatory oversight and enforcement compared to the current framework. Rather than assessing historical compliance breaches against specific obligations, as under the Online Safety Codes and Standards, eSafety will be required to evaluate whether providers have taken appropriate action to identify, assess, and effectively mitigate online safety risks from occurring in the first place.

This will require ongoing investigations, comprehensive evidence gathering, detailed assessments and in some instances extensive scrutiny of a provider's risk assessments and risk mitigation measures over time to establish whether a service has breached their duty.

[3] Delia Rickard PSM (2024) *Independent statutory review of the Online Safety Act*, DITRDSCA website, accessed 5 November 2026.

Even taking a high harm and risk-based prioritisation approach to regulatory efforts, eSafety will need to undertake sustained regulatory engagement to determine whether a prioritised provider's systems and processes are delivering effective safety outcomes in practice and to determine compliance with the digital duty of care. The success of the digital duty of care framework will depend on ensuring eSafety has the powers, capabilities and resources necessary to undertake this complex form of oversight and enforcement.

6. eSafety must be appropriately resourced to support the breadth, scale, and complexity of the amendments proposed under the Bill.

The digital duty of care framework will require more intensive supervision, monitoring, assessment and enforcement activities than the current Online Safety Codes and Standards regime, including oversight of risk assessments, systems-based obligations and broader provider accountability measures. Other amendments in the Bill would also take effect on Royal Assent if passed and have immediate impact on current eSafety resourcing, such as the nudify services removal powers, and amendments to complaint schemes. Adequate and sustained resourcing will be critical to ensure eSafety can effectively perform its new functions, adequately assess risk, enforce compliance, and achieve the policy objectives of the Bill.

For more detail on this, refer to **Appendix D: Implementation considerations.**

7. eSafety strongly supports the inclusion of testing accounts (referred to in the Bill as 'sock puppets') for the sole purpose of performing the Commissioner's functions or exercising the Commissioner's powers under the Online Safety Act.

It is vital that the Online Safety Act provides express authority for eSafety to use such accounts so eSafety can appropriately test, investigate and gather admissible, first-hand evidence with respect to a service's compliance or non-compliance with their legislative obligations. This provision is essential to enable eSafety to verify the availability of certain features on services and test their functionality, to ascertain if a provider is complying with their obligations under the digital duty of care. This is the most practical approach to begin assessing a platform's compliance. Without this coverage eSafety must rely solely on the information and claims provided to it by the platforms themselves, with no mechanism to effectively validate or verify those claims.

eSafety notes the commentary about testing accounts. Testing accounts can only be used for the sole purpose of performing the Commissioner's functions or exercising the Commissioner's powers under the Act.[4] Accordingly, there is no mechanism for testing accounts to be misused, such as to conduct covert operations. The Bill also provides

[4] See Online Safety Act s 27 and proposed s 166B.

mechanisms for public transparency and external accountability. For example, eSafety would be required to prepare annual statements which summarise what the testing accounts were used for the previous year and explicitly set out the priorities and strategies for the year ahead. eSafety would also be required to consult the Commissioner of the Australian Federal Police in determining the policies to follow in using testing accounts.

For further detail on this, see **Appendix E: Testing accounts (sock puppets).**

8. **The age assurance and access control requirements under section 27D apply where there is a reasonably foreseeable risk that a child may access or be exposed to pornography and fetish material. This could be strengthened by expanding the scope of harms. For example, to require age assurance to prevent children's access or exposure to other harms such as high impact sexually explicit material, self-harm and suicide material, high impact violence material and violent instruction material.**

Some services may choose to implement age assurance and access control measures to harms beyond online pornography or fetish material as part of their risk mitigation approach, if they identify a risk of material or conduct harmful to children. If the intention is for age assurance to be a baseline requirement for additional categories of content and conduct, consideration could be given to including further harm types in section 27D. eSafety notes the current Age Restricted Material Codes require age assurance and access control measures for pornography, fetish material, high impact sexually explicit material, self-harm and suicide material, high impact violence material and violent instruction material.

9. **Given the complexities of designing and operationalising the approved researcher access schemes, eSafety suggests it is essential to conduct a scoping exercise once the digital duty of care is in effect to inform the development of legislative instrument establishing the schemes, followed by a staged implementation approach to help identify and address operational, legal and technical challenges before a full roll-out.**

eSafety supports reducing information asymmetries and supporting researchers to gain access to the relevant data to conduct online safety research. Under the Bill, approved researcher access to data schemes may be established under Ministerial legislative instrument.

eSafety notes that implementing a researcher access scheme will be a substantial undertaking necessitating significant additional resourcing and specialised capabilities. Given the complexity of designing and operationalising such a scheme, sufficient time and resourcing is required to ensure it is effective. eSafety suggests it is essential that a scoping exercise is conducted once the digital duty of care is in effect to inform the development of legislative instrument establishing the scheme. This would enable comprehensive consultation with stakeholders, including researchers, privacy and ethics stakeholders, and

industry, and allow the framework to be informed by the practical experiences of jurisdictions that have already implemented comparable models. There are also various matters to determine under legislative instrument including eligibility requirements, approval processes, governance arrangements for approved researchers, data governance, ethical and privacy considerations.

Experience in other jurisdictions demonstrates the scale of this task. For example, the European Union's vetted researcher access scheme took around three years to develop and implement.

For further information about approved researcher access to data schemes, refer to **Appendix F: Approved researcher data access schemes.**

Appendix A: Strengthening the digital duty of care framework

Noting Parliamentary debate and concerns amongst stakeholders regarding the repeal of the Online Safety Codes and Standards, should Parliament wish to provide further specificity in the Bill, the Bill could prescribe specific minimum obligations to comply with the digital duty of care under disallowable legislative instrument. Requirements in legislative instrument could draw on the obligations in the Online Safety Codes and Standards in a streamlined and simplified manner. This would reduce complexities in the existing framework but provide industry with sufficient clarity under the new framework.

eSafety observes that many comparable international frameworks have regulator-developed codes or specific, minimum requirements embedded in legislation alongside systemic obligations such as risk assessments. The Independent Review recommended that eSafety be empowered to make mandatory rules on how entities can comply with the digital duty of care.[5] This would also provide greater clarity to industry in what is expected of them in a manner that is captured under legislation and enforceable, noting that any regulatory guidance eSafety provides is not enforceable.

Case study: International approaches to online safety frameworks

Ireland, *Online Safety and Media Regulation Act 2022* – Outcomes-based protections and prescriptive obligations under regulator-drafted codes

Ireland's legislation empowers its regulator, Coimisiún na Meán, to draft and enforce the legally binding Online Safety Code against a range of designated platforms. The code is split into two requirements, with Part A setting out the outcomes-based protections for users such as obligations for regulated entities to take appropriate measures to protect children and the general public, reinforced by specific obligations in Part B such as requirements regarding terms and conditions, age assurance measures, parental controls and reporting mechanisms. As a European Union member state, Ireland's laws interact with the *Digital Services Act (DSA)*, where some regulated entities are both designated to comply with Ireland's code, and have obligations under the DSA. In these cases, regulated entities must also conduct risk assessments that incorporate the harm categories in the Irish code.

Singapore, *Online Safety (Miscellaneous Amendments) Act 2022* – Regulator-drafted codes

[5] Delia Rickard PSM (2024) *Independent statutory review of the Online Safety Act*, Recommendation 9, DITRDSCA website, accessed 5 November 2026.

Singapore's legislation empowers its regulator, the Infocomm Media Development Authority (**IMDA**), to require regulated services to conduct risk assessments to meet their duties (Section 45L(4ciii)). Under its legislation, IMDA is also empowered to issue legally binding Codes of Practice, which establish specific requirements relating to content moderation, user reporting mechanisms, child-safety protections, and proactive detection of priority harms.

United Kingdom, *Online Safety Act 2023* – Overarching duties, including to undertake risk assessments, complemented by regulator-drafted codes

The UK's legislation imposes multiple 'duties', requiring service providers to identify, mitigate and manage the risks of harm arising from illegal content and activity, and content and activity which is harmful to children. The legislation requires its regulator, Ofcom, to draft codes of practice, with public consultation, about how platforms can meet requirements for specific categories of duties. These codes of practice outline recommended measures that service providers can adopt to demonstrate compliance with the relevant duty. If service providers choose not to implement the recommended measures, they are required to maintain a record that explains how they are otherwise compliant with their duties.[6]

European Commission, *Digital Services Act 2022* – Specific minimum requirements embedded in primary legislation

The *Digital Services Act* imposes tiered, due diligence obligations on digital service providers operating in the EU to identify, assess, and mitigate systemic risks, alongside a range of specific, minimum requirements regarding content moderation and platform accountability. These requirements include content removal and complaints management obligations (Articles 16, 17, 20, 22), algorithm-related obligations including transparency/dark patterns and profiling (Article 25, 27), and specific obligations in relation to the protection of minors (Article 28), risk mitigation (Article 25) and Crisis Response (Article 36). Non-binding guidelines support compliance with these specific rules and requirements.

[6] Note, Ofcom's Codes of Practice provide safe harbours if providers comply with their measures, but are not themselves mandatory, as providers can take alternative measures. The current Bill does not provide for safe harbours - eSafety supports a framework that does not provide safe harbours.

Appendix B: Risk assessments

eSafety supports the inclusion of risk assessments under the digital duty of care. Risk assessments are a key mechanism to operationalise the digital duty of care by providing eSafety with documented information against which eSafety can assess compliance.

eSafety notes the Bill provides the ability to issue remedial directions to services if they fail to conduct risk assessments in accordance with section 27. eSafety would greatly benefit from the clear power to require reassessment, an independent audit, and to seek authorisation for remote viewing powers. This would strengthen eSafety's ability to effectively address inadequate risk assessments, particularly if providers understate key elements of their risk assessment, such as the identified risks, mitigation measures, or measures to assess the effectiveness of mitigation.

eSafety is concerned that services may maliciously comply with the risk assessment framework by providing lengthy risk assessments without meaningful substance, minimising or mischaracterising risk ratings and effectively delaying safety uplift. For example, under the Bill, services are to take into account the likelihood of harm occurring. While this is an important consideration, eSafety is concerned that providers may use prevalence metrics to try to argue that a typical end user's exposure to harms like child sexual exploitation is low in comparison to the sum total of all content available globally. A provider may argue that low prevalence means that it does not constitute a tangible risk on their service and that they do not have to take meaningful action to address it. If this were to occur, providers may reduce their child sexual exploitation detection and prevention capabilities, which are currently required as specific minimum mandatory requirements under the current Online Safety Codes and Standards, and it would be challenging for eSafety to enforce against this reduction in specific safety standards.

Noting the current resourcing of eSafety, we are focused on rigorous prioritisation of legislative obligations and responses to the most harmful material and conduct. Without this change, it is likely that significant additional effort would be expended enforcing compliance with the development and provision of the risk assessment itself rather than the substantive and practical mitigation of risks.

International experience shows that moving from administrative compliance to achieving substantive safety outcomes takes time and can present unique challenges that should be learned from. For example, services may use risk assessment processes to delay safety uplift such as by providing lengthy risk assessments without meaningful substance, minimising or mischaracterising risk ratings, or failing to comply with assessment requirements altogether. This can delay the implementation of safety measures, and hinder regulators' ability to enforce the substantive safety requirements.

Case study: Risk assessments under the United Kingdom's Online Safety Act

In 2025, the UK's online safety regulator, Ofcom, requested and received 104 risk assessment records. Ofcom reported that **dedicated specialist teams systematically reviewed over 10,000 pages** of qualitative and quantitative information with a view to monitoring compliance and considering improvements across industry in how online safety risks are identified and managed.

Ofcom found that while some providers shared comprehensive records, there were **notable issues with many of the records received. Many providers (69/104) failed to provide important information about risk governance, including responsibility for online safety risks**. Ofcom also observed issues with providers' approaches to completing online safety risk assessments. Ofcom found that many records were missing important elements recommended in Ofcom's guidance for illegal content and guidance for services likely to be accessed by children. Ofcom found that providers **were failing to include relevant information needed to demonstrate that they had met requirements.**

Ofcom asked **11 providers to revisit their services' risk assessments**, and 5 providers to reconsider their conclusions about the levels of risk posed by their services. All either shared supplementary information or a revised record, with many amending risk levels as a result of feedback.

Risk assessments legislative instrument

The Bill empowers the Commissioner to make a legislative instrument regarding additional requirements for services to meet to comply with risk assessment requirements under section 27. eSafety envisages that the framework would take a risk-based and proportionate approach, enabling services to tailor their risk assessment processes to the nature and level of risk presented by their service. This could include differentiated expectations for lower-risk services, while requiring additional assessment and mitigation measures where services incorporate higher-risk features, functionalities or design elements. This approach would reduce unnecessary compliance burdens on lower-risk services, and ensure that regulatory obligations are commensurate with the likelihood of harm.

Appendix C: Nudify services

Regulation of nudify services is currently addressed under the Online Safety Codes and Standards. eSafety has taken enforcement action against several of the world's largest nudify services for allegedly failing to comply with their obligations under the Online Safety Act, including failing to implement appropriate safeguards to prevent the creation of synthetic child sexual exploitation material involving Australian school children.

In total, eSafety's engagement and enforcement actions have led to eight of the most widely used nudify services in Australia, either withdrawing access from Australia, or putting in place measures to comply with their obligations under the Online Safety Codes and Standards, such as age-assurance to prevent children's access. These nudify services were collectively receiving over 200,000 visits a month from Australian users.

eSafety notes the Independent Review recommended that Government explore how best to prohibit search engines and app stores from surfacing, selling or distributing nudify services (recommendation 27).[7]

Based on eSafety's experience of how nudify services are accessed and used, section 86B could be strengthened by expanding the scope of services to allow eSafety to give notices across the tech stack, including to infrastructure services such as hosting service providers, internet service providers and pass-through services, requiring them to cease providing services to nudify platforms. As currently drafted, with respect to nudify websites, the removal notice powers would only require search engine services to remove links to the nudify website. While this may reduce the accessibility of the website, the website would still remain available to those who know its URL. Market intelligence shows that many Australian users access nudify websites directly from the URL or via referral websites. Broadening the scope of recipients of removal notices under section 86B would enable more durable disruption of nudify services and ensure that responsibility for addressing these harms is more appropriately shared across the entities that facilitate their operation and availability, strengthening the ability of the framework to achieve its objectives.

[7] Delia Rickard PSM (2024) *Independent statutory review of the Online Safety Act*, DITRDSCA website, accessed 5 November 2026.

Appendix D: Implementation considerations

eSafety is committed to being a future-ready regulator and will leverage its experience gained over 10 years' implementing world-first online safety schemes to implement the digital duty of care within legislated timeframes. A key consideration for implementing and appropriately enforcing the digital duty of care is budget, resourcing and capability. eSafety's priority is to ensure we have the people, skills, tools and operational capacity necessary to enforce the digital duty of care when it comes into effect.

Implementation tasks

The 12-month implementation window will require delivery of a range of complex tasks including but not limited to:

- Establishing and publishing compliance and enforcement priorities
- Drafting legislative instruments, including to establish additional requirements for risk assessments in the first instance and ahead of implementation
- Drafting regulatory guidance on the digital duty of care
- Consultation with a wide range of industry stakeholders to inform eSafety's regulatory guidance and on legislative instruments
- Delivering targeted education and awareness efforts regarding the digital duty of care
- Ongoing enforcement of the Codes and Standards while transitioning to operationalise for enforcement of the digital duty of care framework to ensure Australians are protected online during the transition
- Implementing and enforcing the new nudify services removal scheme
- Updating regulatory guidance for existing eSafety complaints schemes
- Amending processes and procedures for existing complaint schemes

To support implementation, eSafety intends to commence development of regulatory guidance and the legislative instrument on risk assessments as a priority, with the objective of publishing materials regarding risk assessments at least six months before the risk assessment requirements commence. eSafety proposes a public consultation process to inform the development of these important framework documents. This will assist services to prepare for compliance and facilitate a smooth transition to the new framework.

Resourcing and capability

Reform of this scale requires an expansion of eSafety's regulatory capability. The effectiveness of the regime will depend on eSafety having sufficient capacity to implement, oversee, investigate and enforce compliance with the new obligations. Our assessment is that this will require additional resourcing and budget, particularly to support effective enforcement.

eSafety will require additional regulatory, legal and technical resources to operate an effective compliance monitoring and enforcement capability across a broader range of functions. Even adopting a highest harm, highest risk prioritisation approach, eSafety expects the regulatory effort and resourcing required to enforce the digital duty of care, in line with public expectations, will be greater than that required to administer and enforce the Online Safety Codes and Standards. This reflects the complexity of investigating and establishing a breach of the digital duty of care, which requires more ongoing regulatory scrutiny and assessment about whether a provider's risk assessment, controls, governance arrangements, or mitigations were inadequate and where enforcement is dependent on judgments about the appropriateness and effectiveness of a provider's risk management approach and outcomes. By contrast, under the Codes and Standards, eSafety generally commences an investigation and legal proceedings when it has reasonable grounds to determine an obligation has been breached.

To oversee a more proactive regime under the digital duty of care eSafety will need to adopt a rolling schedule of reviewing and analysing risk assessments across a prioritised range of industry sectors to determine whether they adequately assess reasonably foreseeable risks and the appropriateness and effectiveness of mitigating measures. International experience has demonstrated that systems-based regulatory models require significant analytical, technical and investigative capability to assess whether risk management processes are producing meaningful safety outcomes in practice. While eSafety supports risk assessment and transparency requirements as core components of a digital duty of care model, these provisions will need to be designed and implemented with consideration for the regulatory workload and eSafety's available resources. Without sufficient resourcing, regulatory capacity may be largely subsumed processing large volumes of material with limited remaining capacity for substantive compliance and enforcement action.

It is also anticipated a high amount of ongoing industry engagement will be required to field questions and provide support, as well as to engage technical expertise to support eSafety's interpretation of technical documentation around features, functionalities, and processes, including algorithms. Effective regulation in a rapidly evolving technological environment requires more than transparency and self-reported information from regulated services. eSafety will also need specialist technical expertise and tools to strengthen our regulatory visibility and undertake thorough compliance monitoring. This includes access to multidisciplinary expertise, such as data science and adversarial testing, the capacity to conduct or commission independent testing and audits, and access to tools and devices required to assess technologies, risks and platform compliance.

Given the broad obligation, increased complexity and systems-based nature of the proposed regime, it is critical that eSafety is equipped with the necessary powers, specialist

capabilities and sustained resourcing to effectively carry out its functions. Without adequate investment, the effectiveness of the framework will be constrained by regulatory capacity rather than legislative intent. Ensuring eSafety has the capability, tools and resources required to undertake robust oversight and enforcement will be essential to achieving the objectives of the Bill and delivering meaningful improvements in online safety for Australians.

While enforcement options can range from formal warnings to remedial directions to seeking civil penalty orders, court proceedings, judicial reviews and administrative review processes all require a sufficient legal services and litigation budget. It can be expected that regulatory actions will be challenged in court, requiring external counsel and appropriate legal budget so as to not take away from eSafety's core functions/operating budget.

Regulation without enforcement is merely a suggestion. If Parliament intends for Australians to be protected from online harms it must effectively resource the regulator with both staffing and a litigation budget to deliver on enforcement of the digital duty of care obligations.

Operation of existing schemes

During the 12-month transition period, the Online Safety Codes and Standards will remain in effect. During this period, services must still comply with the Online Safety Codes and Standards, and eSafety will continue to undertake investigations and enforcement action so as to ensure there is no gap in the protection of Australians from online harms. Any resources made available at the cessation of the Codes and Standards would be allocated toward enforcement of the digital duty of care, but this does not supplant the resourcing and expertise needed to implement the duty or cover all of its aspects as explored throughout this submission.

Alongside these activities, eSafety will continue to administer our ongoing functions under the Act with existing and growing resource pressures associated with eSafety's administration of complaint and reporting schemes. Complaint volumes continue to grow significantly year on year while resourcing levels have remained comparatively static, placing increasing pressure on existing staff and operational capacity. In FY 2025-26 alone, eSafety received 40,030 complaints, representing a 38% increase on FY 2024-25. Without a commensurate increase in resourcing, eSafety's ability to effectively deliver both its existing responsibilities and new regulatory functions will become increasingly constrained.

Appendix E: Testing accounts (sock puppets)

A sock puppet identity is defined in the Bill as 'a false or fictitious identity assumed by a person in the course of using, or carrying out any activity in relation to, an online service'. Sock puppets can also be referred to as 'testing accounts', 'pseudonymous accounts', or 'research accounts'. The Bill provides that eSafety will be able to use testing accounts for the sole purpose of performing the Commissioner's functions or exercising the Commissioner's powers under the Online Safety Act.

It is vital that the Online Safety Act provides express authority for eSafety to use such accounts so eSafety can test, investigate and gather admissible, first-hand evidence with respect to a service's compliance or non-compliance with their legislative obligations. This provision is essential to enable eSafety to verify the availability of certain features on services and test their functionality, to ascertain if a provider is complying with their obligations under the digital duty of care.

Without legal certainty, it is difficult for eSafety to conduct the basic work of testing and assessing the guardrails that providers state they have employed on their services to comply with their obligations under the Bill, and be certain that any evidence gathered will be accepted by a court. This is the most practical, logical and common sense approach that any member of the public, researcher or investigator would employ to begin assessing how a platform is working. Such information is critical to eSafety's ability to monitor compliance and bring enforcement action against such services. As an example of current limitations in practice, In April 2026, eSafety undertook an exercise to write to 9 providers to request eSafety declared accounts to confirm if content was available to an end-user on the service under complaint scheme investigations. Only 3 providers agreed to eSafety's request, while the remaining 6 either declined or did not respond to eSafety's request.

Without this coverage eSafety must rely solely on information and claims provided to it by the platforms themselves, with no mechanism to effectively validate or verify those claims.

The Bill outlines the specific activities that may be undertaken using testing accounts and provides clear safeguards to support accountability and transparency. The Bill mandates, under Section 205J, that the Commissioner may only assume a sock puppet identity 'for the purposes of performing the Commissioner's functions or exercising the Commissioner's powers under this Act', and any use must be 'only in accordance with section 205K'.

eSafety's regulatory focus is on supporting people experiencing online harm and driving systemic safety improvements by holding platforms to account. eSafety is not a law enforcement agency and does not have a remit under the Online Safety Act to undertake covert operations. eSafety is aware of public commentary that these accounts may be used

to engage with individuals online without disclosing that the account is associated with eSafety. Section 205K only permits the use of testing accounts to engage with a user 'to the extent necessary to prevent closure of an account with the service'. Services often rely on behavioural signals to identify malicious, inauthentic or bot accounts. To avoid a research account being incorrectly identified as such, eSafety may need to undertake limited activity that simulates ordinary account use, including interacting with certain functions of the service. However, it is not eSafety's objective to use test accounts to interact with genuine users; our objective is to ascertain whether services are complying with their obligations under the Act.

The Bill also provides mechanisms for public transparency and external accountability. For example, eSafety will be required to prepare annual statements which summarise what the testing accounts were used for the previous year and explicitly set out the priorities and strategies for the year ahead. eSafety will also be required to consult the Commissioner of the Australian Federal Police in determining the policies to follow when using identities used.

Case study: Examples of how eSafety may use testing accounts

Below are examples of how eSafety would likely use testing accounts to investigate whether services are complying with their obligations under the Online Safety Act, including their duty to ensure a safe online environment.

Online gaming

A provider of an online gaming service states that users playing on a child account cannot interact with an adult account without parental consent. eSafety could create both an adult test account and a child test account on the service and examine whether the service blocks interaction between the two accounts. eSafety would not interact with genuine users in its testing.

Generative AI

eSafety receives reports that a generative AI service is producing pro-terror material, even though the provider's risk assessment says it has guardrails preventing the generation of such material. eSafety could create an account on the service and prompt-test whether the guardrails employed prevent the user from generating this material. If the provider has truly implemented sufficient measures to prevent the generation of pro-terror material, the service should not be capable of generating that material. If the service does not have effective guardrails and illegal material is generated, eSafety would follow procedures to report the incident to relevant authorities (such as the Australian Federal Police) and this would inform further investigation into whether the service is complying with the digital duty of care.

Social media

A provider of a social media service states that its service enables users to switch to a following-only feed, in line with the User Empowerment Tools Determination. eSafety could make a testing account on the service to confirm the option to switch to a following-only feed is available to users, test whether the following-only feed option exists, is readily accessible and that opting out of a recommended feed does not impact the user experience of the service.

Additionally, testing accounts could be used to access parts of a platform that require a user to sign into an account in order to verify the existence, contents and context of social media posts during the review of a complaint made under the Act.

Online pornography

A provider states that it is complying with its obligations to prevent under-18s from accessing its service. eSafety could use a testing account with the age set to under-18 to test whether services have implemented age assurance and access control measures.

Appendix F: Approved researcher data access schemes

The Bill allows for the creation of approved researcher schemes if established under Ministerial legislative instrument. Platforms will continue to hold the most comprehensive data about activity on their services; however, meaningful access for qualified researchers is necessary to independently assess harms, evaluate platform responses, and support evidence-based approaches to improving user safety.

Implementing a researcher access scheme would need to be a phased process. An initial scoping phase would be required to undertake stakeholder consultation, identify suitable governance frameworks, privacy and security safeguards, researcher approval processes and technical infrastructure to ensure data can be accessed and analysed in a safe and responsible way. The scoping exercise would identify resource requirements to inform a future funding request, noting implementing such a scheme would be a substantial undertaking requiring significant additional resourcing and capabilities.[8]

The development of any researcher access to data scheme under the Online Safety Act should also be informed by the practical experience of jurisdictions that have already implemented comparable programs.

Case study: The European Digital Services Act data access and vetted researcher scheme

Article 40 of the Digital Services Act (DSA) makes provisions for vetted researchers to access non-public data from Very Large Online Platforms (VLOPSs) or Very Large Online Search Engines (VLOSEs) for the purposes of studying systemic risk and assessing mitigation measures in the European Union.

The European Commission's Delegated Regulation sets out the procedures for data access under Article 40. Under the scheme researchers submit data access applications to a national Digital Services Coordinator, who verifies the details of the application and then decides whether to pass the data request on to the data provider. The DSA data access

[8] In July 2025, Ofcom provided advice to the UK Government on a potential researcher access scheme under the UK's Online Safety Act. In its advice, Ofcom found that establishing an intermediary for data access could be costly and operationally complex. To be effective it would require additional technical infrastructure, as well as an interdisciplinary team that includes researchers, data scientists, legal experts and governance professionals to oversee access, ensure compliance and foster trust. These demands require sustained investment and careful planning to ensure the scheme achieves its intended aims. Ofcom (2025) *Researchers' access to information from regulated services*, Ofcom website, accessed 5 November 2026.

portal facilitates the applications for data access and ensures the data access process is consistent.

The scheme came into effect in October 2025 (around three years after the DSA came into force) and is still in the early stages of implementation. To date, no research outcomes have been published.[9] Accordingly, the impact, challenges and operational requirements of the scheme are yet to be fully understood.

What is clear at this stage is that there is a significant resourcing requirement associated with its implementation. For example, there are 27 Digital Services Coordinators working together to publish guidance documents and support researcher applications.

eSafety is eager to monitor the first data access requests and research project outcomes under the scheme before considering a similar approach in Australia. The lessons learned from the implementation of this scheme should inform the resourcing, operations and limitations for any research access scheme established in Australia.

[9] European Commission (n.d.) *Research project results*, DSA Data Access Portal, accessed 5 November 2026.