



Australian Government

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

Influence and impacts of social media on Australian Society

SUBMISSION TO THE JOINT SELECT COMMITTEE ON
SOCIAL MEDIA AND AUSTRALIAN SOCIETY

27 June 2024

Contents

Contents	2
Introduction	3
Social media services	3
Role of DITRDCA	4
Existing regulations	5
<i>The Online Safety Act 2021</i>	5
<i>Online Content Scheme</i>	6
<i>Basic Online Safety Expectations</i>	6
Public interest news and journalism support	7
Promotion of media literacy in the community	7
News Media and Digital Platforms Bargaining Code (led by the Treasury)	8
Election integrity assurance measures (led by the Australian Electoral Commission)	8
Policy responses currently being developed	8
The Online Safety Act Review	8
Online Dating Code	9
Age assurance trial	10
Misinformation and Disinformation measures	10
Media reforms	11
National Classification Scheme reforms	11
Dispute resolution measures (co-led with the Treasury)	11
Safe and responsible artificial intelligence (AI) measures (led by the Department of Industry, Science and Resources)	12
Online scams measures (led by the Treasury)	12
Government coordination	13
International regulations	13
European Union (EU)	13
Digital Services Act	13
Digital Markets Act	14
United Kingdom	14
Online Safety Act	14
Canada	15
Online News Act	15
Bill C-63: The Online Harms Act	15
Attachment A	16
Online Safety Act: Complaint and content-based removal notice schemes	16

Introduction

The Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA) welcomes the opportunity to make a submission to the Joint Select Committee on Social Media and Australian Society's inquiry into the influence and impacts of social media on Australian society.

Among a range of other portfolio responsibilities, the Department is responsible for content, broadcasting and classification policy. This submission focuses on matters within the department's portfolio responsibilities.

The department has carriage of policies and programs that address a variety of digital platform services, including social media services, that distribute or facilitate access to news and media content, user-generated content, social interactions, and user-to-user communications. We work closely with the eSafety Commissioner, Australia's online safety regulator, the Australian Communications and Media Authority (ACMA), and other portfolio agencies across government to deliver the Australian Government's policies on preventing online harm and promoting online safety.

This submission provides:

- An outline of regulation of social media administered by the Communications portfolio;
- Policy work underway; and
- the domestic and international context that the government operates in.

Social media services

Digital platform services (also sometimes referred to as digital services, online services or online intermediaries) consist of a range of websites, apps, programs and software that facilitate the exchange of data online. They include social media services; commerce websites, such as marketplaces and financial exchange platforms; business-to-business services; search; streaming; content aggregators; delivery and travel services; app stores; and more.

Social media services are a sub-element of digital platform services that facilitate user-to-user type interactions and/or the sharing of user-generated content. Social media services include purely social networking services, as well as professional networking, dating, chat, messaging, social gaming, and user-generated content-sharing services, among others.

For the purposes of this report, the department has chosen to focus on large multinational social media services where a "feed" of user-generated content is a major feature of the service's user environment. Such services would include Facebook, Instagram, Threads, X (formerly Twitter), TikTok, YouTube, and Snapchat.

In addition to facilitating social connection, social media services enable and support a range of other industries, including news and entertainment, recruitment, commerce, marketing and advertising, and deliver substantial efficiencies for businesses in reaching consumers, and for governments in making services and information available to citizens.

The Asia-Pacific Economic Cooperation's Digital Economy Steering Group states that, "depending on the definition of 'digital economy,' current estimates of the size of the digital economy range from 4.5 percent to 15.5 percent of global GDP. Over the next decade, it is estimated that approximately 70 percent of new value created in the economy will be based on digitally enabled platforms."¹ The Australian Bureau of Statistics has

¹ Asia-Pacific Economic Cooperation (2023). *Digital Economy Steering Group*. Available at: www.apec.org/groups/committee-on-trade-and-investment/digital-economy-steering-group

reported that, in Australia alone, digital activity accounted for 6.1 per cent (\$118.0 billion) of total economic value add (\$1,944.8 billion) in 2020-21.²

On these metrics alone, it is clear that Australians can and do benefit from accessing social media services but the plethora of online harms threaten to significantly undermine these benefits. Social media services generally recognise that exposing users to harms has the potential to significantly undermine user trust and engagement, which has a flow on impact on their business models. In response to this, they adopt a range of measures to try to address online harms.

While social media services are not the origin of many online harms – many have existed well before the internet, let alone the first social media service – the hyper-connectivity created by social media services has the power to significantly amplify them. Amplification includes, for example:

- **More people being exposed than ever before.** This effect can create the perception that exposure to harmful content is ‘normal’, and/or drives widespread commentary across networks, so that it can seem that ‘everyone’ is talking about it. This effect also means that harms are felt across entire communities at once.
- **More exposure for individuals than ever before.** Individuals may be able to process a single exposure to harmful content, but repeated exposure over longer periods can have a compounding or desensitising effect for that person.
- **Increased inability to avoid harmful content.** Individuals may struggle to disengage with, turn away from or tune out social media services, even when it exposes them to harmful content.
- **Increased ease and pace of sharing harmful content.** Sharing information across social media services is both easy and fast, meaning that previously-niche harmful material can very quickly and easily reach mainstream audiences.
- **Making harmful content worse.** As noted above, algorithms often capitalise on attention-grabbing material to capture user engagement. When harmful content becomes ‘normalised’ for users, creators may be incentivised to produce more of such content or to produce more extreme content to elicit the same engagement. Monetisation of such content can create further perverse incentives for both creators and platforms.

All social media service providers have a role to play in preventing their services from being misused to distribute or amplify harmful online materials and behaviours.

Role of DITRDCA

The department leads or closely supports cross-portfolio delivery of a range of existing and developing regulations that address harmful online materials and behaviours, including in relation to administrative oversight of, or policy advice on -

Existing regulations:

- The *Online Safety Act 2021*
- The *Broadcasting Services Act 1992*
- National Classification Scheme
- News Media and Digital Platforms Bargaining Code
- Support for public interest news and journalism
- Promotion of media literacy
- Election integrity measures

Policy responses currently being developed:

² Australian Bureau of Statistics (2022). *Digital activity in the Australian economy, 2020-21*. Available at: www.abs.gov.au/articles/digital-activity-australian-economy-2020-21

- Review of the Online Safety Act
- Online Dating Code
- Age assurance trial
- Combatting Misinformation and Disinformation Bill
- Dispute resolution measures
- Safe and responsible artificial intelligence
- Online scams measures

Existing regulations

The department supports the Australian Government with advice on combatting exposure to harmful online material and ways to promote positive online content and engagement. This includes developing and maintaining legislative and regulatory frameworks to prevent, combat or address online harms.

Safe online engagement is a whole-of-economy issue. As such, the department works closely with agencies across the government to ensure our advice and regulations are relevant, targeted and fit for purpose.

The Online Safety Act 2021

The Online Safety Act commenced in January 2022, and provides the eSafety Commissioner with powers to address specific harms, including cyberbullying of children, cyber abuse of adults, non-consensual sharing of intimate images, and illegal and restricted content. It does this through the establishment of complaint and content-based removal schemes, providing swift, practical assistance to people who have been exposed to harm online (see Attachment A for further detail).

The eSafety Commissioner has powers to investigate complaints and objections made under the Online Safety Act schemes (Part 3 of the Online Safety Act), issue removal notices, and take a range of enforcement actions for non-compliance. The eSafety Commissioner may also issue informal requests to digital platform service providers, which can be effective in quickly removing content without a formal notice.

- **Cyber-bullying of children** - refers to when a child or young person (under 18 years of age) is targeted by online material that is seriously threatening, seriously intimidating, seriously harassing or seriously humiliating. This can include posts, comments, emails, messages, memes, images and videos. The eSafety Commissioner can act on such material across the full range of digital platform services where children spend time, including gaming platforms. Cyber-bullying material targeted at an Australian Child (a child who is ordinarily resident in Australia) is defined in section 6 of the Online Safety Act. Removal notice and formal warning powers are set out in Part 5.
- **Cyber-abuse of adults** - involves persons over 18 years of age and must reach a higher threshold of harm to be acted upon by the eSafety Commissioner. This harmful content meets the threshold if it is intended to cause serious harm, and is menacing, harassing or offensive in all the circumstances. Cyber-abuse material targeted at an Australian adult is defined in section 7 of the Online Safety Act. Removal notice and formal warning powers are set out in Part 7.
- **Non-consensual sharing of intimate images** - intimate images can include images that depict a person's private parts (e.g. genitals, anal area or breasts), engaged in a private act (e.g. undressed, using the toilet, showering, having a bath, or engaged in a sexual activity), or without attire of religious or cultural significance that they would usually wear (e.g. a hijab). In some cases, the image may not be of the victim, but be manipulated to resemble them (e.g. in the case of a 'deep-fake'³), while in other cases the image may be of the individual and either taken unknowingly or without consent to its use. Part 6 of the Online Safety Act addresses the non-consensual sharing of intimate images on digital platform services.

³ A type of digitally-manipulated media (image, video or audio) that replaces one person's likeness with another's. Deep-fakes can range in quality and realism, with the most convincing taking advantage of advanced technologies such as generative AI to make the media appear more real.

OFFICIAL

- **Illegal and restricted content** - refers to a range of harmful online material. It includes material that is illegal or seriously harmful for all audiences, regardless of age – for example, child sexual abuse material or material that promotes, incites or instructs in matters of crime or violence. At the less severe end, it also includes material that is suitable for adults but not children – for example, high impact depictions of sex, nudity, violence or drug use. Removal notice and link deletion powers are set out under the Online Content Scheme in Part 9 of the Online Safety Act.

The Online Safety Act also provides a mechanism for the eSafety Commissioner to request or require the blocking of material that promotes, incites, instructs in or depicts abhorrent violent conduct if the material is likely to cause significant harm to the Australian community. This includes visual or audio-visual material that records or streams abhorrent violent conduct.

- **Abhorrent violent conduct** – includes a person engaging in a terrorist act, murder or attempted murder, or torture, rape or kidnapping of another person. Material that depicts abhorrent violent conduct is defined in section 9 of the Online Safety Act. Powers to request or require material blocking are set out in Part 8.

Under the Online Safety Act, the digital platforms industry is expected to do more to keep its users safe, particularly children. A systems focus is provided through Industry codes and standards established under the Online Content Scheme and through the Basic Online Safety Expectations.

Online Content Scheme

Under the Online Content Scheme, set out in Part 9 of the Online Safety Act, the eSafety Commissioner has powers to remove the ‘worst of the worst’ illegal and harmful content, no matter where it’s hosted. The eSafety Commissioner can assess content for compliance with the Online Content Scheme and, if content is found to be in breach of the Scheme, the eSafety Commissioner can issue a notice to compel digital platform service providers to act. Digital platform service providers that fail to act on a notice within 24 hours can be met with significant financial penalty. Furthermore, the eSafety Commissioner can seek a Federal Court order to remove access to digital platform service providers that continually disregard notices.

The Online Safety Act requires that digital platform service providers develop mandatory and enforceable industry codes to address online harms relating to class 1 (the most seriously harmful online content such as the sexual abuse of children or acts of terrorism) and class 2 (content which is inappropriate for children, such as pornography) materials. If a code does not meet the statutory requirement of providing appropriate community safeguards, the eSafety Commissioner may determine an industry standard. Civil penalties apply for non-compliance with a direction to comply with an industry code or standard.

Basic Online Safety Expectations

The Basic Online Safety Expectations, as set out in Part 4 of the Online Safety Act and determined by the Minister for Communications, set the government’s basic safety expectations for social media services, relevant electronic services and designated internet services. Under the Online Safety Act, the eSafety Commissioner is empowered to require service providers to report on how they are meeting these expectations. The expectations themselves are not enforceable, but non-compliance with the reporting requirements is subject to civil penalties.

On 31 May 2024, the *Online Safety (Basic Online Safety Expectations) Amendment Determination 2024* came into effect. The 2024 Determination includes new expectations that service providers:

- take reasonable steps to ensure that the best interests of the child are considered in the design and operation of their services
- consider user safety and incorporate safety measures into generative AI capabilities and proactively minimise the generation of unlawful or harmful material (such as non-consensual deep-fakes)
- consider safety in the design and operation of recommender algorithms, and proactively minimise the algorithmic amplification of unlawful or harmful material

OFFICIAL

- services take reasonable steps to detect breaches of their terms of use, and respond to user complaints within a reasonable period of time
- make available controls that give users choice and autonomy in deciding who they interact with, the content they receive, and their level of privacy.

Public interest news and journalism support

The government is currently developing the News Media Assistance Program (News MAP). The News MAP is a program of work committed to delivering principles-based and evidence-informed solutions to address the difficult structural challenges facing the sector which prevent news organisations from adequately providing public interest news and journalism.

The department undertook public consultation on a News MAP consultation paper between 12 December 2023 and 22 February 2024. The scope of this consultation has captured feedback from a wide range of news media businesses which are now being assessed by the department. The department will provide advice to the government shortly on the outcomes of the public consultation process.

The government has also implemented measures that support the long-term viability of the public interest news and journalism sector. This includes:

- stable funding for the national broadcasters, including \$6.0 billion for the Australian Broadcasting Corporation (ABC) and \$1.8 billion for the Special Broadcasting Service over 5 years from 2023-24
- \$11 million in 2023-24 and \$12 million in 2024-25 to the Australian Associated Press, to contribute to media diversity by supporting the operation of the AAP newswire service to enable it to continue to service retail news outlets
- \$20 million over 3 years from 2022-23 for the Broadcasting Resilience Program, to improve resilience of ABC sites used for emergency radio broadcasting which are at high risk of service failure due to natural disasters
- \$21.89 million per year (indexed) for the Community Broadcasting Program, to provide funding to community broadcasting organisations to support a range of activities, including transmission and equipment costs and new content development
- \$22.7 million over 4 years from 2023-24 for the Amplifying Australia's Voice in the Pacific program, to strengthen Australia's cultural ties to the region by providing partner Pacific broadcasters with access to a range of Australian television content
- \$0.9 million in 2022-23 and \$0.8 million in 2023-24 to the Public Interest Journalism Initiative, to support provision of its data activities relating to public interest news and journalism in Australia
- \$1.5 million in 2022-23 and 2023-24 to the Local and Independent News Association, to support capacity building for small, hyperlocal news businesses and to collect data on the state of local and hyperlocal news
- \$5 million over 2 years from 2022-23 to the Journalist Fund, to support news businesses to hire new cadet journalists to produce locally relevant core news content in regional areas
- \$15 million in 2022-23 to the Regional and Local Newspaper Publishers Program, to assist print news publishers to absorb newsprint price increases by providing financial assistance including for printing costs, assets and equipment directly related to printing.

Promotion of media literacy in the community

The department delivers the Supporting Media Literacy in Culturally and Linguistically Diverse Communities grant program, which will provide \$1.5 million to the Federation of Ethnic Communities' Councils of Australia (FECCA). FECCA will work with its network to develop a program to distribute small grants and provide guidance, training and support to develop materials and resources that enable communities to critically engage with the information publicised through media channels and be able to spot false narratives, reducing their vulnerability to associated harms.

News Media and Digital Platforms Bargaining Code (led by the Treasury)

The News Media and Digital Platforms Bargaining Code (NMBC) is a mandatory code of conduct that governs commercial relationships between Australian news businesses and ‘designated’ digital platform service providers that benefit from a significant bargaining power imbalance. The NMBC includes minimum standards, including in relation to recognition of original news content; and the incentives created by the designation mechanism, and backed by mandatory negotiation and arbitration, have resulted in substantial payments to support the sustainability of Australian news and journalism.

The impending expiration of Meta’s payment agreements

On 1 March 2024, Meta announced that it would not renew the commercial agreements it made with Australian news media organisations following the introduction of the NMBC, nor would it enter into new agreements.

The government has stated that it is deeply disappointed with Meta’s decision. The government has sought advice from the Treasury and the Australian Competition and Consumer Commission (ACCC) on next steps, and both the government and the department are closely monitoring news content carriage on digital platform services across a number of jurisdictions, including Meta’s actions in Canada.

Election integrity assurance measures (led by the Australian Electoral Commission)

DITRDCA is a member of the Electoral Integrity Assurance Taskforce and Board, both chaired by the Australian Electoral Commission, which were established in 2018, and provide a structured framework for sharing information on potential threats to electoral integrity.

The Taskforce, which includes relevant policy departments as well as a range of national security and law enforcement agencies, facilitates collaboration and engagement across government to protect the integrity of Australia’s electoral processes against threats such as disinformation campaigns, foreign interference, and cyber intrusions.

Policy responses currently being developed

The Online Safety Act Review

On 22 November 2023, the Minister for Communications, the Hon Michelle Rowland MP, announced the commencement of the independent statutory review of the Online Safety Act. As announced in the government’s April 2023 response to the House of Representatives Select Committee on Social Media and Online Safety Report, the government committed to bring forward the review to ensure Australia’s online safety framework remains fit for the changing online environment. The government appointed Ms Delia Rickard PSM to conduct the review and has asked that a report of the review be provided to government by 31 October 2024.

The terms of reference for the review are broad-ranging and include consideration of the overarching objects, operation, and effectiveness of the Online Safety Act, the eSafety Commissioner’s functions and powers, and penalties. The terms of reference also specify that the review should consider whether additional measures are warranted to address online harms not explicitly addressed in the Online Safety Act. Some of the online harms being considered include:

- **Technology-facilitated abuse** – refers to ‘using technology to enable, assist or amplify abuse or coercive control of a person or group of people.’⁴ It can include any form of abuse that is enabled through digital technologies. This includes where technology is used as part of stalking or monitoring, psychological and emotional abuse (including threats), sexual violence or harassment, bullying or hate speech. Specific forms of technology-facilitated abuse include cyber abuse, image-based abuse and technology-facilitated gender-based violence.
- **Volumetric attacks** – refers to when a person is tagged or linked to an abusive post which others like, share, or repost with additional commentary. Often the content is shared with an accelerating level of outrage and toxicity, and ultimately a high volume of abuse. Volumetric attacks often involve abusive posts connected with the target, which others like, share, or repost with additional commentary, and they sometimes involve coordinated and/or disingenuous behaviour.
- **Hate speech** - There are different views about what constitutes hate speech, and it is often highly contested and context dependent. Hate speech generally refers to any form of commentary that expresses hate against a person or a group of persons on the basis of their race, religious affiliation, ethnicity, immigrant status, asylum seeker or refugee status, caste, disability, disease, age, sexual orientation, sex or gender identity. The Online Safety Act does not define online hate or confer specific hate speech-related powers on the eSafety Commissioner, but provides some protections where hate speech overlaps with the material dealt with under its regulatory schemes.
- **Promotion of self-harm material, including eating disorders** - Some online content may deliberately or inadvertently lead to the promotion of disordered eating or other types of self-harm. This type of content can range from deliberate, explicit instructions for how to perform extreme self-harm behaviours and actions, including suicide, to merely suggestive material that might create, encourage or contribute to body dysphoria, such as some types of dieting advertising.

Further, the review is considering the potential harms raised by emerging technologies such as generative AI and recommender systems, as well as international developments in online safety regulation, including whether the law should be amended to impose a new duty of care on digital platform service providers towards their users, and regulatory arrangements ensuring industry acts in the best interests of the child.

On 29 April 2024, an issues paper was released inviting public submissions up until 21 June 2024.

Online Dating Code

On 18 September 2023, the Minister for Communications, the Hon Michelle Rowland MP, wrote to the most popular online dating services in Australia, requesting that they develop a voluntary code of conduct to keep their users safe. The government has asked online dating services to work together to improve engagement with law enforcement, support at-risk users, improve safety policies and practices, and pursue greater transparency about harms. The request has been informed by consultation with the online dating industry, the family, domestic and civil society sector, and policing and civil society.

The code will be led by the industry, with dating platforms responsible for developing its terms. In developing the code, the government has requested that industry consult stakeholders including representatives from the Australian family, domestic and sexual violence sector, victim-survivors, policing, and government agencies.

The digital platforms industry has until 30 June 2024 to develop and adopt its voluntary code. The eSafety Commissioner will evaluate the code after it has operated for 9 months. If the code fails to meet the expectations of government, the Minister for Communications has indicated that legislative options will be pursued to uplift safety for users.

⁴ World Economic Forum (2023). *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms*. Available at: www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf

Age assurance trial

In May 2024, the Australian Government committed \$6.5 million in the 2024-25 Budget to conduct a trial of age assurance technologies. The department will be responsible for the trial. The Prime Minister and the Minister for Communications have acknowledged that the age assurance trial can consider the effectiveness of age assurance as an option for addressing access to pornography by those under the age of 18, and access to social media by children.

This trial will include research, cross-government consultation on assessment criteria, and performance testing of age assurance technologies against these criteria. The trial outcomes will inform policy advice to the government as well as the existing work of the eSafety Commissioner under the Online Safety Act – including through the development of industry codes or standards – to reduce children’s exposure to pornography and other age-restricted services.

Misinformation and Disinformation measures

While misinformation and disinformation are not new phenomena, social media services can have a significant role in enabling or amplifying the spread of false or inaccurate information. This is a serious concern as, regardless of whether mis- or disinformation is unintentional or deliberate, it has the capacity to negatively impact and undermine trust and engagement - from the interactions of individuals, right up to our democracy and society as a whole. Mis- and disinformation can be propagated by online actors, including governments, state backed entities, extremist groups or individuals, and often leverages false identities and anonymous or stolen accounts to seed and/or amplify content through either the actions of individuals or as part of more sophisticated and co-ordinated inauthentic behaviours (CIB).

The aim of spreading mis- and disinformation is to pollute the information environment and sow distrust in government and civil institutions, such as science, journalism and education, and, at its worst, discredit democratic processes. Globally, there are examples of how mis- and disinformation have caused extreme distress among sectors of a society and led to the breakdown of trust within communities, leading to social and public disorder.

Mis- and disinformation have been harnessed to undermine electoral or political processes and this is leading to growing concern among democratic countries about the scale of the problem and its potential impacts. For example, it is expected that mis- and disinformation, particularly that which is proliferated through CIB, will escalate ahead of major elections in 2024. The increased use of generative artificial intelligence (AI) further amplifies the risks and threatens to greatly enhance the scale and reach of CIB campaigns from fine-tuned micro-targeting of individuals and small communities right through to amplified campaigns scaled to a magnitude not previously seen.

The Australian Government has been developing the draft Combatting Mis- and Disinformation Bill to address the growing challenge that mis- and disinformation pose to the safety and wellbeing of Australians, as well as our democracy, society and economy.

The Combatting Mis- and Disinformation Bill would give the ACMA authority to act where industry efforts to combat mis- and disinformation are inadequate. This would include providing ACMA with powers to:

- gather information from digital platform providers on mis- and disinformation
- define rules for digital communications platforms on record keeping, dispute resolution, risk management and media literacy
- register and enforce industry-developed codes of practice covering measures to combat mis- and disinformation on digital platform services
- create and enforce an industry standard (a stronger form of regulation), should a code of practice be deemed ineffective in combatting mis- and disinformation on digital platform services.

The Combatting Mis- and Disinformation Bill would apply to digital communications platform services that are accessible in Australia, such as social media, search engine, instant messaging, news aggregation and

podcasting services, encouraging these organisations to have robust systems and measures in place to address mis- and disinformation on their services that could cause or contribute to serious harm.

The Combatting Mis- and Disinformation Bill would not enable ACMA to request specific content or posts be removed from digital platform services and includes strong protections for privacy and freedom of speech. For example, professional news content, online content related to satire, reasonable public debate (i.e. academic, scientific, religious or artistic), and private messages would be exempted entirely.

Media reforms

The *Broadcasting Services Act 1992* (Broadcasting Services Act) deals with rules around content regulation and media ownership in Australia. However, subject to some exceptions, the Broadcasting Services Act does not generally regulate content provided on internet services.

The current *Broadcasting Services ("Broadcasting Service" Definition—Exclusion) Determination 2022*, and its predecessors, have had the effect of excluding certain types of online media services from regulation under the Broadcasting Services Act, including online television simulcasts, online radio stations, and live-streaming on social media services and other digital platform services, by excluding them from the definition of 'broadcasting service.'

The regulatory delineation between broadcasting and digital platform services, and the potential regulation of the latter, is being considered as part of the government's broader media reform agenda.

National Classification Scheme reforms

Under the National Classification Scheme, all films, computer games and certain publications must be classified under the *Classification (Publications, Films and Computer Games) Act 1995*. The National Classification Scheme is established under a cooperative arrangement with the states and territories through the *Intergovernmental Agreement on Censorship 1995*.

In March 2023, the government announced a two-stage process of reforms to modernise the National Classification Scheme, in part to address regulatory gaps in the online environment. It is not intended that social media services be subject to the requirements of the National Classification Scheme. To this end, the public consultation process on the second classification reforms specifically sought feedback on the potential to carve out the requirement for user-generated content to be classified and regulated under the National Classification Scheme.

Dispute resolution measures (co-led with the Treasury)

DITRDCA works closely with the Department of the Treasury on dispute resolution on digital platforms.

Appropriate and effective complaint and dispute resolution systems and processes are essential for users of social media services. Such systems allow users to report any inappropriate content and behaviours that they encounter on social media services as a first step in prompting further investigation.

On 11 November 2022, the ACCC released its fifth interim report for its Digital Platform Services Inquiry (DPSI), which reiterated the findings and dispute resolution recommendations from the 2019 Digital Platforms Inquiry report, which recommended the government establish mandatory internal dispute resolution (IDR) standards for all digital platform service providers and a digital ombudsman scheme for external dispute resolution (EDR).⁵ The DPSI further clarified that these measures should particularly apply to digital platforms

⁵ Australian Competition and Consumer Commission (2019). *Digital Platforms Inquiry - Final Report*, Page 509. Available at: www.accc.gov.au/about-us/publications/digital-platforms-inquiry-final-report

that provide search, social media, online private messaging, app store, online retail marketplace and digital advertising services.⁶

On 8 December 2023, the government published its response to the DPSI and supported in principle:

- the need to have adequate processes for consumers to raise issues and concerns faced online
- the government undertaking further work to develop IDR and EDR requirements by calling on industry to develop voluntary IDR standards by July 2024.⁷

The Minister for Communications and the Assistant Treasurer implemented this commitment on 6 February 2024, by writing to digital platforms industry representatives and requesting that they develop a voluntary IDR code by July 2024. The digital platforms industry is currently working to deliver a voluntary IDR code for the sector, and the department is closely monitoring industry efforts to progress this work.

Safe and responsible artificial intelligence (AI) measures (led by the Department of Industry, Science and Resources)

The department also provides policy support to the Department of Industry, Science and Resources, which leads the Commonwealth's work on the safe and responsible use of AI.

Through the Safe and Responsible AI agenda, the Australian Government is acting to ensure the design, development and deployment of AI systems in Australia in legitimate, but high-risk settings, is safe and can be relied upon, while ensuring the use of AI in low-risk settings can continue to flourish largely unimpeded.

The government's framework seeks to address these risks including through:

- delivering regulatory clarity and certainty
- supporting and promoting best practice for safety
- ensuring government is an exemplar in the use of AI
- engaging internationally on how to govern AI.

In the 2024-25 Budget the government committed \$39.9 million over five years to develop policy and capability for the use of AI in a safe and responsible manner, including:

- \$21.6 million over four years from 2024–25 to establish a reshaped National AI Centre (NAIC) and an AI advisory body within the Department of Industry, Science and Resources
- \$15.7 million over two years from 2024–25 to support industry analytical capability and coordination of AI policy development, regulation and engagement activities across government, including to review and strengthen existing regulations in the areas of health care, consumer and copyright law
- \$2.6 million over three years from 2024–25 to respond to and mitigate against national security risks related to AI.

Online scams measures (led by the Treasury)

Scams on social media services are a rising issue, with Australians reporting losses of over \$2.74 billion to online scams in 2023.⁸ Some scams on social media services take the form of advertisements for products or services, while others might take the form of comments or messages sent directly to users. Of particular concern are the increasing instances of images of Australian public figures being non-consensually used to promote investment and other types of financial scams.

⁶ Australian Competition and Consumer Commission (2022). *Digital Platforms Services Inquiry – September 2022 interim report*, Page 88. Available at: www.accc.gov.au/about-us/publications/serial-publications/digital-platform-services-inquiry-2020-25-reports/digital-platform-services-inquiry-september-2022-interim-report-regulatory-reform

⁷ Australian Government (2023), *Government's response to the ACCC Digital Platform Services Inquiry*. Available at: <https://treasury.gov.au/publication/p2023-474029>

⁸ Australian Competition and Consumer Commission (2024). *Targeting Scams Report 2023*. Available at: www.accc.gov.au/system/files/targeting-scams-report-activity-2023.pdf

The government is implementing a range of measures to combat online scams, including:

- establishing the National Anti-Scam Centre, which builds on the works of the ACCC's Scamwatch service, to coordinate government, law enforcement and the private sector to combat scams
- announcing an intent to introduce legislation establishing mandatory industry scams codes on banks, telecommunications providers, and social media, digital messaging and search advertising services.

The department is supporting the Treasury in its coordination of implementation of the government's scam commitments, including by leading work to address telecommunications scams and by providing input on dispute resolution measures for mandatory industry codes.

Government coordination

The department works closely with our counterparts across government and provides secretariat support for the Online Harms Ministers Meeting (OHMM).

The government established the OHMM to better coordinate online harms policy, in response to the March 2022 Report from the House of Representatives Select Committee on Social Media and Online Safety. The OHMM meets bi-annually and is Chaired by the Minister for Communications.

The OHMM brings together Commonwealth Ministers with responsibility for addressing online harms to discuss cross-cutting issues and coordinate policy across portfolios. OHMM members are Ministers with responsibility for regulating digital platforms, as well as those who oversee portfolios where the impact of online harms are felt.

At their meeting on 4 March 2024, OHMM members discussed a range of cross-cutting issues, including the role that digital platform services can play in undermining gender equality, and efforts to address harmful gender stereotypes promoted online. Ministers also discussed work to address hate speech on digital platform services and the government's efforts to combat scams.

International regulations

Many international jurisdictions are similarly concerned with the online harms stemming from social media services. A number of countries have taken steps to regulate a range of online harms, with the European Union, United Kingdom and Canada in particular leading the way. The department continues to monitor international developments and draw on these where relevant to shape and support Australia's digital platform regulatory reforms.

European Union (EU)

In 2022, the EU adopted the Digital Services Act package, made up of the Digital Services Act (DSA) and the Digital Markets Act (DMA).

Digital Services Act

The DSA is targeted at online intermediaries and platforms, including online marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms, with additional, stricter obligations for very large online platforms and search engines (online platforms and intermediaries that have more than 45 million users per month in the EU). The DSA ensures:

- an easy way to report illegal content, goods, or services
- stronger protections for people targeted by online harassment and bullying
- transparency around advertising

- bans on certain types of targeted advertising, such as those using sensitive data or the data of minors
- easy-to-use, free-of-charge complaint mechanisms for if an online platform takes our content down
- simplified terms and conditions.

Digital Markets Act

The DMA is targeted towards designated 'gatekeeper' platforms. Gatekeeper platforms are digital platforms with a systemic role in the internal market that function as bottlenecks between businesses and consumers for important online services, such as for example online search engines, app stores, messenger services.

The DMA requires gatekeepers to:

- allow third parties to inter-operate with the gatekeeper's own services in certain specific situations
- allow their business users to access the data that they generate in their use of the gatekeeper's platform
- provide companies advertising on their platform with the tools and information necessary for advertisers and publishers to carry out their own independent verification of their advertisements hosted by the gatekeeper
- allow their business users to promote their offer and conclude contracts with their customers outside the gatekeeper's platform.

The DMA additionally prohibits gatekeepers from:

- treating services and products offered by the gatekeeper itself more favourably in ranking than similar services or products offered by third parties on the gatekeeper's platform
- preventing consumers from linking up to businesses outside their platforms
- preventing users from un-installing any pre-installed software or app if they wish so
- tracking end users outside of the gatekeepers' core platform service for the purpose of targeted advertising, without effective consent having been granted.

On 6 September 2023, the European Commission designated six digital platform providers as gatekeepers under the DMA. Alphabet, Amazon, Byte Dance, Meta and Microsoft were required to make their 22 core platform services compliant with the DMA by March 2024.⁹

United Kingdom

Online Safety Act

The UK Online Safety Act addresses illegal content and content harmful to children online. The Act became law in October 2023, and establishes:

- the Office of Communications (Ofcom) as the regulator for online safety, with appropriate enforcement, oversight and designation responsibilities
- requirements to scan for child sexual exploitation and abuse material, and to report such material to the UK National Crime Agency
- and duty of care obligations, requiring platform providers to act against a range of illegal and harmful online material, including:
 - false communications intended to cause non-trivial psychological or physical harm
 - threatening communications, including threats of death, serious injury, rape, assault by penetration, or serious financial loss
 - flashing images intended to trigger a seizure, alarm or distress for a person with epilepsy
 - communications encouraging or assisting serious self-harm or suicide

⁹ European Commission (2023). *Digital Markets Act: Commission designates six gatekeepers*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4328

- cyberflashing communications
- sharing or threatening to share intimate images or film without consent.

Canada

Online News Act

Similar to the News Media and Digital Platforms Mandatory Bargaining Code, the Online News Act in Canada establishes a framework for digital news intermediary operators and news businesses to enter into agreements to make available news content on digital platform services.¹⁰ The Act received Royal Assent on 22 June 2023.

Obligations under the Online News Act will come into effect no later than 180 days after Royal Assent. When elements of the Act come into effect will depend on regulations from the Governor in Council, and the implementation of processes by the Canadian Radio-television and Telecommunications Commission (CRTC). The CRTC is consulting publicly on the regulatory framework.

Bill C-63: The Online Harms Act

Bill C-63: Online Harms Act of Canada proposes to introduce a number of reforms aimed at reducing online harms, including:

- requirements for online platforms to monitor and remove seven categories of harmful content, which broadly relate to non-consensual intimate content, content that victimizes children, and content that foments hatred or incites violence
- the establishment of three new regulatory bodies, including:
 - a Digital Safety Commission, to monitor and order removal of content
 - a Digital Safety Ombudsman, to focus on systemic issues relating to online content moderation
 - a Digital Safety Office of Canada, to support the Digital Safety Commission and the Digital Safety Ombudsperson in fulfilling their mandates.
- new offences under the Criminal Code of Canada, to address hate speech content and messaging
- new powers to impose cost recovery charges on social media service providers, to fund the operation of the regulations.

¹⁰ Parliament of Canada (2023). *Statutes of Canada 2023, Chapter 23 – An Act respecting online communications platforms that make news content available to persons in Canada*. Available at: <https://laws-lois.justice.gc.ca/eng/annualstatutes/index2023.html>

Attachment A

Online Safety Act: Complaint and content-based removal notice schemes

Table 1 – Overview of complaint and content-based removal notice schemes under the Online Safety Act

	Image-Based Abuse	Child Cyberbullying	Adult Cyber-Abuse	Illegal and Restricted Content
Online harm	Posting or threatening to post an intimate image depicting another person without that person's consent (irrespective of whether the image has been altered).	Online material that is likely intended to have an effect on an Australian child, and likely to have the effect of seriously threatening, seriously intimidating, seriously harassing or seriously humiliating the Australian child.	Online material that is likely intended to have the effect of causing serious harm to an Australian adult and would reasonably be regarded as menacing, harassing or offensive in all the circumstances.	Online material that is Class 1 or Class 2 material (determined by reference to Australia's National Classification Code).
Who is protected?	Person depicted (or purported to be depicted).	A targeted child (who is ordinarily resident in Australia).	A targeted adult who is ordinarily resident in Australia.	End-users in Australia.
Link required to Australia	The person depicted, or the person who posted or threatened to post, is ordinarily resident in Australia (or, for objection notices only, the image is hosted in Australia).	Material is targeted at a child ordinarily resident in Australia ('Australian child').	Material is targeted at an adult ordinarily resident in Australia ('Australian adult').	Material suspected to be accessible to Australians online. Class 1 material can be hosted anywhere, but Class 2 material must be provided by a service in Australia or hosted in Australia.
Who can make a complaint?	The person who has reason to believe an intimate image depicting them has been shared without consent (or that a threat to share such an image has been made); a person authorised by the depicted person; or a parent or guardian of the depicted person.	The targeted Australian child or a parent, guardian or responsible person authorised by the child or an adult who was an Australian child.	The targeted Australian adult or responsible person authorised by the Australian adult.	A person who resides in Australia, or an entity that carries out activities in Australia, or an Australian Government. (Note, eSafety can investigate material within this scheme without receiving a complaint).
Does complainant need to report to the service provider before a removal notice can be issued?	No	Yes	Yes	No (can be reported anonymously)

Table 2 – eSafety Commissioner’s complaint scheme compliance and enforcement powers

	Image-Based Abuse	Child Cyberbullying	Adult Cyber-Abuse	Online Content Scheme
Formal warning to person who posts or threatens to post image	Yes	No	No	No
Removal notice to service provider/hosting service provider	Yes	Yes	Yes	Yes
Removal notice to end-user	Yes (a ‘removal notice’)	Yes	Yes (a ‘removal notice’)	No
Remedial direction to end-user	Yes	No	No	No
Remedial notice to service provider	No	No	No	Yes (Class 2 only)
Service provider notification	Yes	Yes	Yes	No
Service provider statement	Yes	Yes	Yes	Yes
Link deletion notice	No	No	No	Yes (Class 1 only)
App removal notice	No	No	No	Yes (Class 1 only)
Federal Court order to cease providing service	No	No	No	Yes (in exceptional situations)
Alternative enforcement arrangements	Formal warnings, enforceable undertakings, court injunctions, infringement notices, civil penalty orders and financial penalties.			

These compliance and enforcement powers involve the following:

- **Removal notice:** Notice requiring recipient (end-user, service provider or hosting service provider) to remove material or stop hosting material within 24 hours or longer period the Commissioner allows (civil penalty for non-compliance).
- **End-user notice:** Notice requiring person who posted cyberbullying material targeted at a child ordinarily resident in Australia to: remove the material and/or refrain from posting cyber-bullying material targeting the child and/or apologise for posting the material (enforceable by injunction).
- **Remedial direction:** Direction to end-user who has posted or threatened to post intimate images without consent to take specified remedial action to prevent future contraventions (civil penalty for non-compliance).
- **Remedial notice:** Notice requiring the recipient to remove Class 2 material or to make the material subject to a Restricted Access System (civil penalty for non-compliance). The eSafety Commissioner may declare by legislative instrument that a specified access control system is a restricted access system under section 108 of the Online Safety Act.