

**Submission to the Consultation on the Second Action Plan
(National Plan to End Violence Against Women and
Children 2022-2032)**

July 2026

Executive summary

Technology-facilitated abuse (TFA) is among the most pervasive, rapidly evolving, and stubbornly gendered harms confronting Australian women and girls. It spans a broad spectrum—including image-based abuse, sexual extortion, online harassment, surveillance, financial abuse, and technologically enabled coercive control.

Its burden falls overwhelmingly on women: almost 80% of eSafety non-consensual intimate image matters (excluding sexual extortion) involve female victims, 94% of recorded coercive control incidents target women, and 97% of intimate-partner homicides reviewed in New South Wales involved coercive control. TFA is no longer a marginal online irritation; it is increasingly the primary medium through which domestic and gendered violence is planned, rehearsed, and executed.

Despite its prevalence, TFA lacks an explicit, named place in Australia's regulatory architecture. It remains scattered across criminal statutes, legacy classification law, and voluntary codes. Even the Government's proposed Digital Duty of Care framework folds TFA into a disconnected array of Criminal Code cross-references rather than recognising it as a coherent harm in its own right.

The result is a fragmented ecosystem where regulators lack a conceptual anchor, high-risk offenders exploit cross-platform gaps, and insidious, longitudinal harms like coercive control fall through the cracks. Drawing on direct experience in regulatory design and law enforcement, Primary Purpose Advisory (PPA) advances a single organising proposition for the Second Action Plan: Australia must give TFA an explicit, statutory identity and build targeted, intelligence-led enforcement on existing national infrastructure rather than inventing new institutions.

Building an Intelligence-Led Regulatory Response

TFA offending is inherently cross-platform. Perpetrators fluidly move across social media, encrypted messaging, and dating apps to target victims, ensuring no single platform holds a complete picture of the offending. To break this playbook, Australia should adapt its mature financial-sector threat-intelligence model (such as APRA's CPS 234) to online safety.

Using the existing capability of the Australian Federal Police-led Joint Policing Cyber Crime Centre (JPC3), the Government should establish a national TFA threat-intelligence capability and an Adverse Actor Register. This would enable closed-loop, privacy-preserving signaling across digital platforms, financial institutions, and courts to disrupt recidivist accounts and coordinate action against high-risk offenders—backed by strict, domestic-violence-informed safeguards to prevent the misidentification or weaponisation of the system against victims.



Harnessing Digital Platforms as Sentinels and Diversion Points

The embedded nature of digital technology means platforms are uniquely positioned to serve as early-warning "sentinels" against coercive control and potential femicide. Rather than creating new capabilities, platforms should be required to analyse existing behavioural and network signals—such as anomalous social graph constriction or shared device credentials—to detect patterns of coercive control. Dedicated research by the AIC, ANROWS, and eSafety should establish a Coercive Control Digital Signals model to determine how platforms can safely support early intervention.

In parallel, the demonstrated power of platform-level warning and diversion interventions should be deployed across search, social, and messaging services to guide women seeking help and to divert men concerned about their own coercive behaviours.

Countering Attitudinal Conditioning, Incel Ideology, and AI Harms

The cultural substrate driving male violence is adapting rapidly online. Misogynistic incel ideology has evolved into mainstream subcultures like "looksmaxxing," which repackages toxic "blackpill" nihilism for platforms like TikTok. Concurrently, the rapid rise of uncensored AI companion chatbots presents a critical new frontier of attitudinal harm.

Designed around compliance and sycophancy, these platforms allow boys and young men to rehearse scripts of male dominance and non-consensual subjugation with submissive, female-coded avatars. AI companion platforms must be treated as a distinct harm category requiring dedicated regulatory oversight beyond content classification, focusing on architectural harm, emotional manipulation, and strict age assurance.

Structural Repairs to the Online Safety Architecture

These proactive capabilities must be matched by real structural repairs to existing legislation.

First, the eSafety Commissioner's frontline TFA Support Service faces an imminent funding cliff and requires sustainable, ongoing Budget allocation.

Second, the Online Safety Act 2021 must be decoupled from the outdated National Classification Scheme—which assesses real-world violence against women through commercial film criteria—and rebuilt around a statutory taxonomy of harms.

Third, the Adult Cyber Abuse scheme requires root-and-branch reform to address its unworkable thresholds and account for cross-platform "volumetric" attacks.

Finally, and in recognition of its role as a vector for financial scams and fraud, the fast-growing online dating sector must be moved from an unenforceable voluntary code

onto a mandatory statutory footing, with its current exemption under the Scams Prevention Framework removed.

Anchoring Reform in a National Digital Safety Strategy

Australia's current approach to digital safety is deeply fragmented across at least six Commonwealth portfolios, leaving technology-facilitated abuse owned by every department in theory, but none in practice.

Just as Australia built a national strategy and dedicated coordinator to defend its digital systems against cyber threats, it must now build an equivalent blueprint to defend its people.

The Second Action Plan should advocate for a cross-portfolio National Digital Safety Strategy which is anchored in a shared taxonomy of digital harms and overseen by a designated national mechanism. The Strategy should integrate responsibilities across Communications, Attorney-General's, Home Affairs, Treasury, Industry, and Social Services. Only through a unified national strategy can the digital dimensions of violence against women be met with the focus, authority, and systemic rigor they demand.

Contents

Executive summary	2
Introduction	6
About Primary Purpose Advisory	6
Summary of recommendations	7
Technology-facilitated abuse	9
Assistance for victims of TFA	12
Recognition of the cross-platform nature of TFA.....	13
Adverse actor threat intelligence.....	15
Could the Register harm the women it aims to protect?	18
Coercive control patterns detection.....	19
Does this amount to a new national surveillance capability?	22
Attitudinal issues: interventions and emerging risks	24
Warnings, deterrence, and help-seeking diversion	24
Behaviours and attitudes of men and boys	25
From Tate to bonesmashing to blackpill nihilism	25
The new frontier of attitudinal conditioning: AI chatbots	29
Why uncensored chatbots could do more harm over time than pornography	31
Other systemic issues	32
The Zarutska Board decision as a classification failure	34
Adult cyber abuse – not fit for purpose	35
The Digital Duty of Care – flawed on TFA	37
Online Safety Code for Dating Services	38
Tracy Hall’s story	39
Towards a national digital safety strategy	42

Introduction

About Primary Purpose Advisory

Primary Purpose Advisory (PPA) is a specialist consultancy advising governments, regulators, law enforcement agencies, civil society organisations, and digital platforms on online safety framework design, child safeguarding, platform accountability, and regulatory strategy.

PPA was founded by [REDACTED], a senior regulatory and law enforcement leader with over two decades of experience operating at the intersection of public safety, child protection, and digital regulation.

[REDACTED] began his career as a Detective with the NSW Police Force investigating major and organised crime, before playing an operational leadership role in New South Wales child protection reforms under *Keep Them Safe*, establishing pioneering diversion services for vulnerable youth. In 2013, he joined the Australian Communications and Media Authority to lead frontline operations disrupting online child sexual abuse material (CSAM) and extremist content in close partnership with global law enforcement networks.

With the establishment of the eSafety Commissioner in 2015, [REDACTED] played a central, hands-on role in shaping Australia's national online safety architecture. Over a decade at eSafety, he served as inaugural Executive Manager of the Investigations Branch, Chief Operating Officer, and ultimately General Manager of the Regulatory Operations Division at the Senior Executive Service (Band 2) level. In these capacities, [REDACTED] oversaw the expansion of Australia's regulatory scope from CSAM to cyberbullying, image-based abuse, sexual extortion, terrorist content, and emerging digital harms, while leading operational implementation of the landmark Online Safety Act 2021.

Internationally, [REDACTED] has served as Vice President of the International Association of Internet Hotlines (INHOPE) and represented Australia before global bodies including the WeProtect Global Alliance, INTERPOL, the European Commission, and the OECD. He has also advised the private sector, including through appointment to Westpac's *Safer Children, Safer Communities* expert roundtable. He holds academic degrees in law, international security studies, and politics.

This submission is directly informed by [REDACTED] operational grounding, statutory enforcement experience, and strategic expertise in translating complex regulatory, investigative, and policy challenges into practical, defensible solutions that protect the public interest.

Summary of recommendations

A national TFA taxonomy. That a comprehensive taxonomy of TFA harms be created to guide policy-making and the funded measures delivered through the Second Action Plan.

Sustainable funding for the eSafety Commissioner's TFA Support Service. That ongoing funding for the eSafety Commissioner's TFA Support Service be included in any National Plan action package, at least matched to the current one-year, non-ongoing allocation of \$5.4 million.

An Adverse Actor Register through the JPC3. That the existing capability of the AFP's Joint Policing Cyber Crime Centre be used to create a national TFA intelligence program, requiring in-scope digital platforms to remove high-risk user accounts through an Adverse Actor Register, enabled by appropriate powers and legislative reform — for example, a scheme equivalent to APRA's CPS 234.

A Coercive Control Digital Signals model. That dedicated research (for example, jointly through the AIC, ANROWS and the eSafety Commissioner) be undertaken to map TFA-related coercive behaviours to digital platforms and produce a Coercive Control Digital Signals model; and that this research assess whether the proper, most effective and safest role for platforms is as a longitudinal early-warning mechanism or as a provider of preserved data as retrospective evidence for law-enforcement operations.

Platform-level help-seeking diversion. That, in consultation with the digital platforms industry, the Government explore how online services can support help-seeking through targeted diversion interventions, both for women experiencing coercive control and for men concerned about their own behaviour.

Research into the evolution of incel ideology. That the Government concentrate national research efforts on tracing the continued evolution of incel ideology into forms such as looksmaxxing, in recognition of the serious risk of harm posed to women by "blackpill" nihilism.

Dedicated regulation of AI companion platforms. That AI companion platforms be treated as a distinct harm category requiring dedicated regulatory attention beyond content classification and general privacy frameworks. These AI apps require binding safety standards across safety, privacy and security domains that address emotional-manipulation architecture, dependency-inducing design, and lack of effective age assurance.

Root-and-branch review of the Act's adult-harms logic. That the Government, with respect to the Online Safety Act 2021, consider root-and-branch reform of the Act's adult harassment, abuse and intimidation logic, including whether the Adult Cyber

Abuse scheme is fit for purpose. Government should also commit to legislating against cross-platform "volumetric" attacks.

Decouple the Online Safety Act from the Classification Scheme and explicitly name TFA in any successor harms taxonomy. That the Online Safety Act be decoupled from the National Classification Scheme, as recommended by the Rickard review, and that TFA be given its own explicit, named place in any successor taxonomy – including within the anticipated Digital Duty of Care.

Enforceable, industry-wide regulation of online dating services. That the Government legislate the Online Safety Code for Dating Services and, more broadly, undertake reform to (a) define dating apps as a regulated sector under the Online Safety Act 2021; (b) ensure future codes and/or the Digital Duty of Care explicitly cover dating apps and related harms, including financial abuse; and (c) amend the Competition and Consumer (Scams Prevention Framework—Regulated Sectors) Designation 2026 to remove the exclusions currently afforded to dating apps.

A cross-portfolio national digital safety strategy. That the Government develop a comprehensive national online safety strategy spanning the Communications, Attorney-General's, Social Services, Home Affairs and Industry portfolios, deliberately intersecting with existing strategies including those on child sexual abuse and violence against women.

Technology-facilitated abuse

Technology-facilitated abuse (TFA) is a form of harm that is multifarious, rapidly evolving, and escalating.

As a broad label, TFA includes image-based abuse, sexual extortion, harassment and intimidation online, scams and financial abuse, surveillance and monitoring, online stalking, and coercive control behaviours such as denying access to online accounts.

It is well-established that TFA is highly gendered, with most forms perpetrated against women and girls. In addition to the figures cited in the Consultation Paper, eSafety Commissioner data shows that, once sexual extortion is excluded, almost 80% of all matters involving threatened or actual sharing of intimate material without consent – including within a coercive control context – involved women.¹ A similar proportion of eSafety complaints were made by women where intimate material was posted to public-facing websites without consent.² In one of the largest categories of image-based abuse, peer-group sharing, the most heavily targeted group is girls aged 13-15 years.³

Even when sexual extortion is considered, the gender differences are stark. While men aged 18-24 comprise the largest group in the eSafety data targeted by sexual extortion, with almost 80% of all reports involving male victims,⁴ offender preferences show distinct patterns.

Men and boys are typically targeted for financial gain, with offenders using social engineering techniques to rapidly leverage shame, guilt and fear for monetary benefit.⁵ In contrast, girls and women tend to be targeted for *content*, with demands centred on coercion of more intimate material over funds.⁶ Female victims are more likely to be younger than men and boys at the time of first victimisation (14.6 years vs 15.4 years), are more likely to be revictimised,⁷ and will likely be abused more often by someone they know.⁸

In relation to other forms of TFA, women are more than 20% more likely to be victims of dating app-facilitated sexual violence than men.⁹ Girls are at far higher risk of being

¹ Burton et al, 'Regulating Image-Based Abuse: An Examination of Australia's Reporting and Removal Scheme', *Journal of Online Trust and Safety*, April 2025, 16.

² Ibid.

³ Ibid, 17.

⁴ Ibid, 15.

⁵ Wolbers H et al, 'Sexual extortion of Australian adolescents: Results from a national survey', Trends and Issues in Crime and Criminal Justice No. 712 2025 < https://www.aic.gov.au/sites/default/files/2025-02/ti712_sexual_extortion_of_australian_adolescents.pdf > 16.

⁶ Ibid.

⁷ Ibid, 18.

⁸ Ibid, 21.

⁹ Wolbers H and Dowling C, 'Routine online activities and vulnerability to dating app facilitated sexual violence', Trends and Issues in Crime and Criminal Justice No. 704 2024

targeted for online sexual solicitation from adults than boys (26.3% vs 7.6%) and will generally experience both the non-consensual sharing of intimate images online *and* sexual solicitation from adults at a higher rate (6.5% vs 0.8%).¹⁰ One in four girls will be approached by an adult online for sexual material before age 12.¹¹

Finally, women experiencing domestic, family and sexual violence are at greater risk from related online and technical harms. Almost all frontline practitioners surveyed for the 2020 WESNET *Second National Survey on Technology Abuse and Domestic Violence in Australia* stated that their clients had experienced technology-facilitated stalking and abuse.¹² Respondents indicated that 93% of TFA victims in their work were female, and males account for 96% of abusers.¹³ Social media services are widely and commonly used as the vector for abuse and threats made towards women, as well as surveillance.¹⁴

The perpetration of TFA within family and domestic violence (FDV) contexts keeps pace with technological advances, with offenders using commercial ‘off the shelf’ tools to track, surveil, and monitor targets – tools often collectively known as ‘spyware’.¹⁵ In 2022, the Australian Federal Police charged a Brisbane man with creating and selling a ‘remote access tool’ (RAT) dubbed Imminent Monitor to 14,500 customers in 128 countries.

The agency told *The Guardian* that a ‘statistically high’ number of Australian purchasers were named on domestic violence registers.¹⁶ New technologies, such as ‘smart cars’, wearables and ‘Internet of Things’ devices are increasingly being used to perpetrate abuse.¹⁷ In 2024, the NSW Crime Commissioner commenced ‘Project Hakea’ to investigate the use of tracking and surveillance devices to enable serious crime,

<[https://www.aic.gov.au/sites/default/files/2024-](https://www.aic.gov.au/sites/default/files/2024-11/ti704_routine_online_activities_and_vulnerability_to_dafsv.pdf)

11/ti704_routine_online_activities_and_vulnerability_to_dafsv.pdf> 6.

¹⁰ Walsh K, Matthew B et al, ‘Prevalence and characteristics of online child sexual victimization: Findings from the Australian Child Maltreatment Study’, *Child Abuse and Neglect* 160 (2025), 5.

¹¹ Ibid.

¹² Woodcock D et al, *Second National Survey on Technology Abuse and Domestic Violence in Australia* 2020 <<https://wesnet.org.au/wp-content/uploads/sites/3/2020/11/Wesnet-2020-2nd-National-Survey-Report-72pp-A4-FINAL.pdf>> 2.

¹³ Ibid, 13.

¹⁴ Ibid, 22-24.

¹⁵ Australian Strategic Policy Institute, ‘Digitally watched without consent: spyware as a tool of coercive control’ 18 May 2026 <<https://www.aspistrategist.org.au/digitally-watched-without-consent-spyware-as-a-tool-of-coercive-control>>.

¹⁶ The Guardian, ‘Brisbane teenager built spyware used by domestic violence perpetrators across world, police allege’ 30 July 2022

¹⁷ eSafety Commissioner, ‘From smart cars to tracking devices: Technology’s increasing role in coercive control and family and domestic violence’ 30 November 2025

<<https://www.esafety.gov.au/newsroom/blogs/from-smart-cars-to-tracking-devices-technologys-increasing-role-in-coercive-control-and-family-and-domestic-violence>>.

including domestic violence, after finding that one-quarter of tracking devices sold to NSW customers were bought by DV offenders.¹⁸

Coercive control behaviours involving online technologies are common and becoming normalised. Recently, the NSW Bureau of Crime Statistics and Research found that well over half of all coercive control incidents recorded by NSW Police between July 2024 and December 2025 involved harassment, monitoring or tracking (56%); threats or intimidation (53%); financial abuse (52%) or these in combination along with nine other behaviour types.¹⁹ Almost all victims in these incidents were women (94%),²⁰ and on average four coercive controlling behaviours were recorded per incident.²¹

In its 2025 report on tech-based coercive control attitudes, the eSafety Commissioner found that indisputably coercive behaviours were substantially normalised by younger Australians – especially men. Surveying a representative sample of more than 2,000 Australians, eSafety found that men were far more likely to agree that explicitly coercive behaviours (such as constantly texting to check what a partner is doing or who they are with) were signs of care within an intimate relationship.²²

Concerningly, more than 10% of surveyed men agreed that care was shown by insisting that checking on a partner *and* expecting them to be constantly available to respond to text messages *and* checking how a partner looks in a photo before they post it.²³ In response to whether tracking a partner wherever they went using digital services such as Life360 or Find My Friends was a reasonable expectation within a relationship, nearly one in five respondents aged 18-24 and 35-44 agreed.²⁴

The analytical snapshot provided above is not intended to be exhaustive, but merely to demonstrate the protean and evolving nature of TFA, and its vastly disproportionate impact on women and girls. Despite its diverse spectrum, TFA can be systematised within a categorisation schema. There is enough research and evidence obtained from victim-survivors to be able to produce a comprehensive taxonomy of TFA harms that reflects the Australian experience. This taxonomy should inform the way TFA is recognised by and addressed through the Second Action Plan.

¹⁸ NSW Crime Commission, 'Media Release: NSW Crime Commission investigation into the criminal use of tracking and other surveillance devices' 25 June 2024 <<https://www.crimecommission.nsw.gov.au/media/2024-media-statements/hakea-media-release-25-june-2024.pdf/@download/file>>.

¹⁹ NSW Bureau of Crime Statistics and Research, 'Coercive control monitoring report: December 2025' <https://bocsar.nsw.gov.au/documents/topic-areas/domestic-violence/Coercive_control_monitoring_report_Dec_2025.pdf> 1.

²⁰ Ibid, 5.

²¹ Ibid.

²² Office of the eSafety Commissioner, *Track, Harass, Repeat: Attitudes that normalise coercive control 2025* <<https://www.esafety.gov.au/sites/default/files/2025-04/Track-harass-repeat-attitudes-around-tech-based-coercive-control.pdf?v=1785131551060>> 11.

²³ Ibid, 10.

²⁴ Ibid, 16.

Recommendation 1: A comprehensive taxonomy of TFA harms should be created to guide policy-making and funded measures achieved through the Second Action Plan.

Assistance for victims of TFA

The significance of TFA as a growing source of harm affecting women and girls is becoming more apparent as connected technologies are further embedded into daily life. Accordingly, awareness of TFA's many faces and its impacts has improved, leading to various reforms being enacted in the last five years that specifically recognise discrete harms.

These include Commonwealth reforms such as the creation of the *Online Safety Act 2021* (Cth), which enhanced the existing powers of the eSafety Commissioner relating to image-based abuse complaints. Consequential amendments made at the same time created the aggravated Commonwealth criminal offence of using a carriage service to transmit sexual material without consent, providing federal criminal sanctions for image-based abuse.²⁵

All Australian jurisdictions have now criminalised image-based abuse. The relevant statutory provisions typically target the recording and sharing of material along with threats. New South Wales, for example, provides for penalties of three years' imprisonment where a person films, distributes, or threatens to send explicit content without consent. The state goes further, by explicitly prohibiting the creation of 'deepfake' material involving altered or synthetic material.²⁶

Both New South Wales and Queensland have amended state criminal legislation to explicitly recognise coercive control, and in particular technical means of abuse. To illustrate, NSW explicitly provides for monitoring and tracking behaviours, including via technology, within the definition of 'abusive behaviours' within Division 6A of the Crimes Act. In his Second Reading speech introducing the related Bill, then-Attorney General Mark Speakman explained that the reforms were intended to move beyond existing stalking, intimidation and harassment offences to address digital forms as a key tactic of control and isolation.²⁷

However, while these reforms are commendable, there are very limited avenues of support for those experiencing TFA where civil regulatory or criminal options are unavailable or undesirable.

²⁵ *Criminal Code Act 1995* (Cth), s 474.17A.

²⁶ *Crimes Act 1900* (NSW), ss 91P, 91 PA, 91Q, 91R.

²⁷ Parliament of NSW, *Legislative Assembly Hansard*, 12 October 2022 <<https://www.parliament.nsw.gov.au/parliamentary-business/hansard/hansard-full-details?id=HANSARD-1323879322-127842§ion=HANSARD-1323879322-127845>> 9048.

Since 2024, the eSafety Commissioner has provided a TFA Support Service to assist frontline workers in their support of clients experiencing TFA. To provide the service, eSafety has built a call centre capability and has partnered with cybersecurity experts to give tailored advice addressing client-specific TFA concerns. This advice is augmented by a range of resources such as TFA guides for frontline staff, and comprehensive audit checklists.

The service was originally funded in the 2022-23 Federal Budget, which allocated \$16.6 million over four years for its establishment. Originally intended to terminate in 2025-26, Government allocated an additional \$5.4 million over a single year in the last Budget. The measure is now due to terminate in 2026-27. Government is yet to commit to supporting the service on an ongoing basis through the Budget process.

The service fills a critical gap, nationally. Since its inception, it has directly assisted more than 600 TFA victims who otherwise would be hard-pressed to find technically-competent advice from other sources. By partnering with cybersecurity experts, and regularly consulting the DFV sector on its operations, the TFA Support Service is able to maintain currency and legitimacy as a provider of real support to augment existing case management methodologies. Unless the funding cliff is addressed in a sustainable way, not only to provide current baseline services but to meet increasing demand, genuine and urgent needs will be unmet.

Recommendation 2: Include ongoing funding for the eSafety Commissioner's TFA Support Service through any National Plan action package. Funding should be at least matched to the current one-year non-ongoing Budget allocation of \$5.4 million.

Recognition of the cross-platform nature of TFA

The perpetration of TFA frequently involves a wide variety of online platforms and technologies – often within the one incident. Half of young Australians surveyed about online sexual extortion reported that they were asked to switch platforms through the course of offending against them.²⁸ The survey makes explicit what sexual extortion investigators have long observed – a fluid and permeable approach to offending that ranges across online services most commonly used by Australian teens, including Snapchat, Instagram, and WhatsApp.²⁹

It is common for offenders to make first contact with a target on one service (for example, Instagram) to develop a degree of affinity and rapport, before shifting to another (such as Snapchat or an encrypted messaging service) where exchange of intimate material occurs and the 'trap' snaps shut. Recently, the eSafety Commissioner noticed a new entrant to this list – Tinder – suggesting a concerning shift in offender

²⁸ Wolbers H et al (n 5) 13.

²⁹ Ibid.

methodology away from a predominant focus on social media to including dating platforms in the ‘killchain’.³⁰

Other forms of TFA frequently involve multimodal offending methods, especially within the context of DFV. In addition to perpetrating harassment and abuse over social media and messaging services, DFV offenders may employ hidden cameras, easily concealed tracking devices (such as AirTags) and abuse of phone location apps (such as ‘Find My’) to create a powerful digital ‘leash’.³¹ Manipulation of smart and ‘Internet of Things’ (IoT) devices is increasingly common to create fear and distress, for example remotely locking and unlocking doors, changing heating settings and triggering house alarms.³²

What this summary shows is the way that TFA offenders – including within DFV contexts – distribute offending over multiple online platforms and modalities. These include digital services such as social media and encrypted messaging services, consumer-facing services such as those provided by banks and other providers, and devices such as trackers, smart technology, and IoT devices.

*The effect of this for many victims of TFA is a sense of suffocating ‘omnipresence’, where their offender (whether or not known to them) feels to be everywhere, all at once.*³³

Focusing on the role played by social media and instant messaging services in the perpetration of TFA shows opportunities for change. In addition to the desire to achieve blanket monitoring and harassment, one of the effects brought about by offenders using accounts held across multiple services is that no one company holds a complete picture of offending behaviour. If an offender makes contact with a victim on Service A, moves the conversation to Service B, then posts image-based abuse material to Service C, only fragments will be known by each company – which in isolation may not be considered to breach relevant terms of service.

Similarly, when an offender’s account is removed for breaching community standards or guidelines, a move from the enforcing platform to a new platform will not raise risks or flags. In many cases, it is possible for the recidivist offender to simply create a new account on the enforcing platform, given a very low risk of detection. Fluidly moving

³⁰ Office of the eSafety Commissioner, ‘Periodic Notice 3: A snapshot of eSafety’s findings’ <<https://www.esafety.gov.au/industry/basic-online-safety-expectations/child-sexual-exploitation-and-abuse-material-and-activity/periodic-notice-report-3-snapshot>>.

³¹ Woodcock D et al (n 12), 19.

³² Rogers, Michaela M et al, ‘Technology-Facilitated Abuse in Intimate Relationships: A Scoping Review’ *Trauma, Violence and Abuse* Vol 24(4) 2023, 2218.

³³Ibid.

between platforms, creating multiple accounts, and reviving identities on services through recidivist accounts are behaviours that are part of the standard TFA playbook.

Adverse actor threat intelligence

Successful evasion of platform policies and habitual abuse of the mainstream social media and messaging ecosystem could be made more difficult if companies shared threat intelligence about ‘bad users’ with others in the ecosystem. Such a model could be based on existing and well-established financial or cyber threat intelligence models providing for intelligence sharing and threat-preparedness information between Australian regulators and financial institutions, for example. Ensuring resilience is a key principle underpinning the cyber defensive activities of Australian banks, which demands a detailed understanding of threat actors and the overall threat environment.

Australia's financial sector cybersecurity architecture operates through a trusted intermediary model centred on the Australian Cyber Security Centre (ACSC). Financial institutions share threat data upward to the ACSC, which sanitises and abstracts it into actionable indicators. These include malicious IPs, malware signatures and behavioural patterns, which are then distributed back to industry participants. No institution shares raw data laterally with competitors – the ACSC acts as the neutral hub. This is underpinned by the Australian Prudential Regulatory Authority's CPS 234 mandatory reporting obligations, which compel regulated entities to participate – it is not a voluntary scheme. The result is a sector-wide risk-pooling architecture where the privacy and competitive sensitivity problems of bilateral sharing are solved.

To create an analogous architecture for the sharing by members of the online ecosystem is less burdensome than it may appear on its face. Already, the Australian Federal Police-led Joint Policing Cyber Crime Centre (JPC3) functions as a sophisticated joint capability to combat high volume cybercrime. Through memorandums of understanding and joint agency agreements, the JPC3 includes members of state and territory police services, several major financial institutions, AUSTRAC, the Australian Criminal Intelligence Commission, the eSafety Commissioner and the National Anti-Scam Centre.³⁴

The JPC3 would be a natural home for any TFA-specific threat intelligence capability, eliminating the need to create any new national agency or clearinghouse. eSafety intelligence holdings could be merged and cross-referenced with financial and law enforcement agency (LEA) intelligence, and policy enforcement logs of bad user accounts provided to the JPC3 by online service providers (such as social media services and online dating services) above a particular size. Once verified, an ‘Adverse

³⁴ Australian Federal Police, *AFP Submission to the Parliamentary Joint Committee on Law Enforcement: Inquiry into the capability of law enforcement to respond to cybercrime*, 22 February 2023 <<https://www.afp.gov.au/DocumentStore.ashx?id=e3d5cb01-38ea-4099-a4eb-c7fff19f0a88&subId=752009>> 15.

Actor Register’ signal could be sent by the JPC3 to the digital ecosystem, providing a centralised method of coordinating actions against bad user accounts.

Where digital platforms hold information about offending behaviours and patterns of Australian accountholders, provision could be made to share that information in a secure and privacy-preserving fashion with Australian banks and other critical service providers to consider whether, for example, debanking action should be commenced.

Over time, and subject to amendments made to relevant statutory disclosure provisions within Commonwealth legislation, other stakeholders could be included as both sources and ‘customers’ of this kind of Adverse Actor intelligence.

In this list, the Federal Court of Australia (FCA) would feature prominently, given their rich data holdings about DV offending and their intelligence-led approach of identifying and managing high risk cases through the Lighthouse Program.³⁵ Noting the high degree of sensitivity attached to Federal Court risk data, and with all appropriate safeguards in place, using Lighthouse holdings to enrich intelligence in support of the Adverse Actor Register could have a powerful preventive impact.

The desired end state would be a secure, closed ecosystem of key nodes within the TFA prevention and protection space sharing information bidirectionally to support early, swift and intelligence-informed action to address TFA online.

A number of other threshold questions naturally occur, including the size of services within scope of the program, security of signals, and legislative constraints.

In relation to the size of relevant services, the existing ‘Tier 1’ risk assessment definitions contained within the online industry codes of practice provide some guidance, especially with respect to functionality, size, and purpose.³⁶ Additional criteria that might be considered could include harms prevalence data relevant to the service and market share considerations.

Security concerns could be ameliorated by building a cryptographic signalling layer that uses symbolic account indicators instead of personally identifiable information such as account names or user handles. A similar cryptographic schema might be appropriate to handle highly sensitive risk data supplied by the Federal Court.³⁷

There are several legislative hurdles to overcome. The first is building a regulatory architecture mandating involvement of in-scope digital platforms. Such a framework

³⁵ See the FCA website for information about the Lighthouse Program <https://www.fcfa.gov.au/fl/fv/lighthouse>.

³⁶ Various industry organisations, *Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material)* < https://onlinesafety.org.au/wp-content/uploads/2023/06/230616_1_SMS-Schedule_REGISTERED-160623.pdf > 5.

³⁷ By ‘cryptographic signalling layer’, we refer to the practice of substituting ‘plaintext’ (such as personally identifiable information) with ‘ciphertext’, encrypting the former in a way that is still uniquely identifiable but not capable of revealing an identity. CLAUDE PLEASE CHECK THIS FOR ACCURACY.

could be based on APRA's CPS 234, which mandates among other things that a regulated entity "must maintain an information security capability commensurate with the size and extent of threats to its information assets, and which enables the continued sound operation of the entity".³⁸

An equivalent standard could be connected to a register of entities regulated by the eSafety Commissioner at the Tier 1 risk level, and might require the entity to "maintain an Adverse Actor response capability commensurate with the size and extent of threats to its Australian user base, allowing the *safe* operation of the entity".

The most logical home for legislative provisions related to this proposal would be the *Online Safety Act 2021* (Cth). Amending the eSafety Commissioner's existing responsibilities under s 27 may be wise, reflecting the program's data governance requirements. In addition, a statutory limitation on purpose and use of data obtained in service of the program would prevent scope creep and provide avenues for merits and judicial review where data was employed outside the program's boundaries.

Additionally, amendments may need to be considered to the *Privacy Act 1988* (Cth), given the demanding threshold required under the Australian Privacy Principles for personal information to be shared without consent.³⁹ This threshold question could be bypassed with a carveout provided in the Act allowing for certain cryptographic signals to be shared between the AFP and members of industry for a defined purpose.

Statutory disclosure powers provided in Commonwealth legislation would need to be carefully reviewed. Presently, for example, the *Online Safety Act* provides relatively few situations where information can be disclosed by the eSafety Commissioner to specific entities without consent. A list of authorities is provided at s 212 to which the eSafety Commissioner can disclose information, if the information will enable or assist the authority to perform or exercise any of the authority's functions or powers. These naturally include other digital regulators (but not all), and law enforcement agencies.

However, the eSafety Commissioner has no express power to disclose information to the Federal Court (to assist, for example, with the FCA's risk assessment process), nor to agencies or organisations working in frontline DV support. Outside of the usual complaint-management and investigative processes under the Act, the eSafety Commissioner has no power to disclose data to members of the Australian digital industry, such as social media services.

Finally, such a program would need to be underpinned by effective appeal and redress avenues. In addition to administrative law avenues, these avenues should include

³⁸ Australian Prudential Regulation Authority, 'CPS 234 Information Security' <<https://www.apra.gov.au/standards/cps-234>>.

³⁹ The Australian Privacy Principles permit the sharing of personal information without consent in certain exceptional circumstances, if there is a reasonable belief that the disclosure is necessary to lessen or prevent a serious threat to the life, health, or safety of an individual or the public.

existing ombuds functions (such as those overseeing the banking sector), and internal review mechanisms. Powers could be provided to the eSafety Commissioner or another independent regulator to review contested decisions to remove Adverse Actor accounts by digital platforms, and direct reinstatement.

Could the Register harm the women it aims to protect?

Two versions of this objection must be taken seriously.

The first is deliberate weaponisation: that an offender could try to turn the Register against a current or former partner through false or malicious reports, for example to a platform, to the eSafety Commissioner, or to police. An offender might calculate that a confected "adverse" identity could see her account removed through an Adverse Actor signal, thus deepening her isolation.

The second is systemic misidentification, where the Register flags the wrong person. It is well established that domestic and family violence is frequently misread by the systems that respond to it. Perpetrators often present as calm, credible and cooperative, while victims who have resisted or reached breaking point present as volatile, or even aggressors. The result is that women are misidentified as the predominant aggressor in a meaningful proportion of cases.⁴⁰ This is especially true for First Nations women, migrant and refugee women, and women living with disability.⁴¹ A system such as the Adverse Actor Register that ingests complaints, reports and enforcement logs is at risk of perpetuating misidentification if it blindly transmits without verification.

The first objection is answered through the structure of the JPC3, which would act as a clearinghouse. No Adverse Actor signal would issue until the underlying intelligence had been tested through standard quality-assurance processes establishing source reliability and information credibility. To corrupt the Register, an offender would need to manipulate multiple independent systems at once, each with its own verification and integrity controls — a structurally improbable feat.

An answer to the second objection lies in systems design. It requires a stringent set of safeguards to be implemented, because misidentification is not improbable but routine.

PPA recommends that the following safeguards be treated as non-negotiable features of the Register:

- **Corroboration over single sources.** A signal should never issue based on the strength of one complaint or one log. It should require convergent and tested

⁴⁰ Reeves, Ellen 'The continuing problem of misidentification for family violence victim-survivors' (Monash Gender and Family Violence Prevention Centre) < <https://lens.monash.edu/the-continuing-problem-of-misidentification-for-family-violence-victim-survivors>>.

⁴¹ Ibid.

evidence from multiple independent sources, so that neither a malicious report nor a single misidentification can, by itself, trigger.

- **A domestic-violence-informed clearinghouse.** The JPC3's assessment should apply a specialist, gender-informed lens calibrated to known misidentification patterns to enrich standard intelligence QA. This may help identify records that fit the signature of a victim being misread as an aggressor.
- **A victim-protective check.** Where the subject of a proposed signal is, or has been, a protected person in any linked system (such as the National Domestic Violence Order Scheme) that fact should trigger heightened scrutiny.
- **A human, DFV-competent decision-maker should be cemented into the loop before any action is taken.** No account removal, and certainly no debanking, should follow automatically *solely* from a signal.
- **Disparate-impact auditing.** The Register's outputs should be monitored on an ongoing basis and disaggregated by sex. This is so that anomalous, unexpected or disproportionate flagging of women can be detected early.
- **Confidential by design.** The fact that a person has been considered for, or placed on, the Register should never be disclosable to another party, so the process cannot itself become a tactic of abuse. Amendments to Freedom of Information and Federal Court legislation may be required.
- **Fast, assisted reinstatement.** Redress must assume that a misidentified victim may be in crisis and unable to navigate an ordinary appeal, and provide a supported, expedited pathway to reversal.

Recommendation 3: Use the existing capability of the AFP's Joint Policing Cyber Crime Centre to create a national TFA intelligence program, requiring digital platforms to remove high-risk user accounts through creation of an Adverse Actor Register, enabled by appropriate power and legislative reform for example through a scheme equivalent to APRA's CPS 234.

Coercive control patterns detection

Before turning to this proposal, it is worth distinguishing it clearly from the Adverse Actor Register discussed above, because the two operate in different directions. The Register directs detection outward, at the accounts of those perpetrating abuse, in order to disrupt offending. The proposal that follows directs detection toward identifying women who may be at risk, in order to enable protection.

Each raises different questions and require different safeguards. The first focuses on offenders, and so needs to ensure the risk of false positives misidentifying victims as perpetrators is minimised, or even eliminated. The second is a protective mechanism

intended to protect victims. Safeguards here need to operate in a way that does not interfere with their autonomy, or expose them to greater danger.

As noted above, coercive control exercised online and connected devices is a common feature of DFV perpetrated against women. Where dedicated coercive control legislation has been enacted in Australia, such as in NSW, the definition of abusive behaviour explicitly includes tech-facilitated abuse methods, including stalking, surveillance and monitoring.

Coercive control is deliberately and structurally practised as a form of behaviour that is subtle, difficult to detect, and layered. It is often deployed in a way that causes victims to doubt their own senses and instincts, with offenders using ‘gaslighting’ techniques to dismiss, belittle or mock victims’ experiences of abuse. Offenders ‘fly under the radar’, using techniques or tactics that may appear harmless in isolation, but terrifying when rendered as patterns of abuse and control.⁴²

To fully capture coercive control as a strategy of violence and abuse, systems and institutions need to be better calibrated to contribute to pattern identification. The need for bold, creative ideas about how to achieve this could not be more urgent or serious. The NSW Domestic Violence Death Review Team found that, for all intimate partner domestic violence homicides between 2000 and 2022, 97% involved some form of emotional and psychological coercive control.⁴³ In more than a quarter of those cases, there was no known history of physical violence.⁴⁴

In the context of abuse of digital platforms, it may be useful to think of certain coercive control behaviours and their consequences as discrete signals. Collection and analysis of these signals over time can reveal patterns. As noted in the literature, coercive control will generally only be rendered visible when the patterns are.⁴⁵ While the pattern’s full tapestry may only be fully uncovered by stitching together what many *institutions* know about a victim (such as the courts, police forces, health services, and others), the embedded nature of online technologies mean that digital platforms could be recruited as key partners to assist with flagging and understanding potential risks of coercive control, and as a corollary, femicide – digital ‘sentinels’.

Digital platforms such as social media and online dating services are uniquely placed to play a role in enabling intelligence-led interventions by law enforcement and other public authorities. This role would be predicated on timely detection of behavioural and

⁴² Di Marco, Martin Hernan, and Evans, Dabney P, ‘Flying under and through the radar: Tactics used by intimate partner femicide perpetrators to evade interventions’ *International Journal of Law, Crime and Justice* 82 2025, 1.

⁴³ NSW Government, ‘What are the impacts of coercive control?’ <<https://www.nsw.gov.au/family-and-relationships/coercive-control/what-are-impacts>>.

⁴⁴ Ibid.

⁴⁵ Eriksson L, Mazerole P, and McPhedran S, ‘Giving voice to the silenced victims: A qualitative study of intimate partner femicide’, *Trends and Issues in Crime and Criminal Justice* No. 645 2022, 5.

other signals that together suggest on a reasonable basis that an Australian woman is the victim of coercive control. While the exact signals, thresholds and notification pathways relevant to this role require more research, the gravity of Australia's femicide catastrophe – with *four* women and girls murdered in a *single week* in early July 2026 –⁴⁶ suggests that better and more comprehensive early warning mechanisms are urgently needed.

An indicative list of the kinds of signals that might, *when taken together*, point to coercive control includes:⁴⁷

- Statistically anomalous reductions in the size of a victim's social graph – A major coercive control tactic is social isolation, often involving the complete constriction of a victim's access to her social group (although isolation may not be so overt). Offenders who deploy this tactic may seek to replace their partner's social group with theirs, sometimes (but not necessarily) with the complicity of their own male peer circle.⁴⁸ If a victim's *graph topology* (i.e., a user's connection to other users) contracts in ways that are statistically unlikely through organic means, and a partner's network is grafted onto the victim's, this could be a sign of coercive isolation.
- A victim's account profile showing evidence of regular or high-frequency logins from devices associated with another user of the service (inferred as being a partner based on the social graph). eSafety's research into coercive control shows that almost 30% of men expect their partner to provide passwords or login details to accounts.⁴⁹
- Inconsistent or incongruous victim user behavioural signals, suggesting that another person (for example, a coercive intimate partner) is in control of the account.
- Behaviours and activity providing insight into the 'leaking' of offender's thoughts, intentions, fantasies, and emotions. Research suggests that interception of such leaks is a promising strategy for early warnings and opportunities for intervention.⁵⁰ Where other signals suggest that a partner-user of a service is perpetrating coercive control, leaked sentiments could provide powerful additional signals of risk.
- Victim behaviours and activity which show help-seeking, attempts to create trusted (and likely out of security and necessity, covert) connections with family

⁴⁶ Melissa McKay, 'Safety of women and girls under national spotlight again after four alleged murders in four days' *ABC News* < <https://www.abc.net.au/news/2026-07-14/domestic-violence-national-spotlight/106911176>>.

⁴⁷ Noting the need for further research on how TFA-enabled coercive control occurs, especially over social media platforms.

⁴⁸ DiMarco and Evans (n 38), 7.

⁴⁹ eSafety (n 22), 16.

⁵⁰ DiMarco and Evans (n 38), 7./

and friends, and attempts to obtain information from sources about domestic violence and coercive control. In some cases, victims disclose that they understand that a relationship is controlling or that they are being gaslit, even if they may not understand the full implications.⁵¹ Where platforms are capable of seeing messages to this effect, such disclosures may be key signals to cohere pattern analysis.⁵²

- If the eSafety Commissioner is empowered through the JPC3 proposal outlined above to share signals with digital platforms, such signals could be useful augmentation to pattern identification – even more so if other institutional data is included.

Such an early warning system could function to give effect to one of the specific recommendations established in the Commonwealth’s National Principles:

“Responses should also consider different types of behaviour used by perpetrators, including those that are subtle and insidious, and mechanisms through which behaviour can be facilitated, including abuse of systems and use of technology.”⁵³

Does this amount to a new national surveillance capability?

This proposal raises an uncomfortable potential risk, and it is fair to question whether it would turn digital platforms into instruments of surveillance over the private lives of the women it aims to protect. However, the premise of the question is fragile.

The proposal does not ask platforms to build any new surveillance capability. In fact, the signals it describes (including network topology, login and device patterns, behavioural anomalies, message content and sentiment) are not data platforms would need to begin collecting. They *already* capture all of it, and far more. Of that data, platforms already subject it to inference more sophisticated than anything contemplated here – all in the service of advertising and engagement.

Two risks appear to be most prominent.

The first is autonomy. The model detects signals about a woman, and potentially acts on them, without her knowledge. But a foundational principle of domestic and family violence practice is that the victim is the expert in her own safety, with safety planning a delicate, individual, and victim-led action. Not every woman experiencing coercive

⁵¹ Ibid, 6.

⁵² There is a far broader information ecosystem needed to fully capture and identify harmful coercive patterns. This section assumes a degree of help-seeking via digital channels, however deactivating social media accounts altogether is not uncommon when victim-survivors feel vulnerable and exposed to monitoring or tracking. Although beyond the scope of this submission, when thinking about the adjacent systems victim-survivors are likely to access for help in those situations, such as 1800 RESPECT, ensuring that data is captured and appropriately employed to assist pattern detection will be essential.

⁵³ Attorney-General’s Department, *National Principles to Address Coercive Control in Family and Domestic Violence* <<https://www.ag.gov.au/system/files/2023-09/national-principles-to-address-coercive-control-family-and-domestic-violence.PDF>> 31.

control wants intervention, and unsolicited "help" delivered over her head can be disempowering – dangerous if her accounts are being monitored or even used by an offender.

The second is escalation. The period around detection and separation is the most dangerous in an abusive relationship, and the point at which intimate-partner homicide is most likely. The model itself assumes an offender may be monitoring or controlling the victim's account. It follows that any response visible to that offender could be read as a provocation and trigger retaliation. Design choices need to ensure that the system's assistive responses do not generate a fresh signal visible to the offender.

These risks are precisely why this submission recommends research to determine the model's safest form, rather than immediate deployment. They also point to the design principles that research should treat as non-negotiable:

- **Resource, not report, as the default response.** The safest intervention is almost always to surface support and choice to a victim-survivor. That connects this model to the help-seeking diversion approach at Recommendation 5. Such an approach is preferable to notifying a third party without awareness and consent. Notification of law enforcement or another authority should be reserved for narrowly defined circumstances of imminent, serious risk, on a domestic-violence-informed threshold.
- **Responses must be invisible to a monitoring offender.** Where support is surfaced, it must be delivered through channels and at times that do not create a signal an offender could detect.
- **Co-design with victim-survivors and with vulnerable communities.** The model must be built with the DFV sector and, critically, at risk and vulnerable women.
- **Strict use-limitation, absolute prohibitions on repurposing, and independent oversight.** Signals should be tuned to, and only ever used for, the defined protective purpose. Data should never be repurposed for enforcement against the woman herself, fed back into commercial profiling, or disclosed to the offender.

Recommendation 4:

- That specific research be undertaken (perhaps jointly through AIC, ANROWS and the eSafety Commissioner) to map TFA-related coercive behaviours and activities to digital platforms in order to produce a Coercive Control Digital Signals model.
- The research should include an assessment of whether the proper, most effective, and safest role for digital platforms in this scheme is as an early warning mechanism based on collation and assessment of longitudinal data, or as a provider of preserved data as retrospective evidence to support LEA operations.

Attitudinal issues: interventions and emerging risks

Warnings, deterrence, and help-seeking diversion

Consistent with the models outlined above, digital platforms can play a crucial role in providing tailored, in-the-moment information about coercive control to both victims and offenders, and provide links to support services.

It is reassuring to see that Google Search (via the google.com.au domain) provides a link to 1800-RESPECT when users search for phrases such as ‘I think I am a victim of coercive control’ or ‘I am concerned about my behaviours in relation to coercive control’.

However, the degree to which other digital platforms such as social media services produce similar interventions is unclear.

It should be uncontroversial to expect digital platforms to play a role in enabling help-seeking for coercive control victims. Research shows that women experiencing coercive control do try to find information and assistance via digital channels. In fact, Eckstein has shown a correlation between the *level* of coercive control and psychological victimisation experienced and the *intensity of use* by victims of digital channels to seek help.⁵⁴ Noting the ways by which offenders work to systematically isolate women from their social circles, this may be a residual effect: women experiencing intense coercive control end up seeking help via digital channels *precisely because they are the only channels left available*.

Platform-level interventions, for example messages supporting help-seeking, can have powerful effects on user behaviour. In 2026, researchers examined warning messages deployed in a large-scale field study centred on a dark web search engine. During a 70-day campaign, users seeking out child sexual abuse material (CSAM) using the search engine were presented with one of ten random warning messages. Each message had a

⁵⁴ Eckstein, Jessica J, ‘To Harm and Heal: Digital Disclosure and Support-Seeking for IPV Victims’, *Open Information Science* (5) 2021, 185.

unique framing, but all provided a link to divert users away from the search to an online anti-CSAM self-help program. The results were remarkable: click-through rates to help resources doubled over the campaign period, compared to prior.⁵⁵

These findings mirror earlier Australian research, which signalled the potential for warnings to have significant effects on user behaviour, specifically deterrence. That study, by Prichard and others, was aimed at testing popup warnings deterring access to a site simulating CSAM. Results showed a significant reduction in attempts to click through to material users were led to believe may be illegal when shown warning messages about illegality.⁵⁶

This research supports a dual conclusion. Given the extent to which women in coercive control situations seek help online, and the efficacy of diversionary interventions at the platform level, digital platforms should be expected to play a greater role in helping at-risk users to find timely advice and support. However, the same logic and method is likely to be equally helpful to men who are worried about their own coercive behaviours, and seeking help to change. Helpfully, Prichard et al have now published a framework for deploying warning messages over a more extensive slice of the ‘tech stack’, borrowing cybersecurity principles of ‘defence in depth’.⁵⁷

Recommendation 5: In consultation with the digital platforms industry, the Government should explore how online services can play a role in supporting help-seeking through targeted diversion interventions – both for women experiencing coercive control, and for men concerned about their own behaviours. Given that a quarter of fatal domestic incidents were not preceded by recorded violence, reaching women experiencing purely coercive control could help to meaningfully shift help-seeking rates.

Behaviours and attitudes of men and boys

From Tate to bonesmashing to blackpill nihilism

There has been much said and written in recent years about the concerning attitudes towards women and girls exhibited by boys and younger men. Worth specifically noting is the eSafety Commissioner’s joint research with Deakin University and QUT published

⁵⁵ O Ciardha, C et al, ‘Warning Message Content Increases Help Seeking in a Large-Scale Dark Web CSAM Intervention’ arXiv pre-print 2026 <<https://arxiv.org/pdf/2606.06417v1>> 16.

⁵⁶ Prichard J et al, ‘Effects of automated messages on Internet users attempting to access ‘barely legal’ Pornography’ 2021 <https://www.researchgate.net/publication/351609229_Effects_of_Automated_Messages_on_Internet_Users_Attempting_to_Access_Barely_Legal_Pornography/link/62ba9e045e258e67e10b031a/download?tp=eyJjb250ZXh0ljp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19> 20.

⁵⁷ See Prichard, J et al, ‘How to implement online warnings to prevent the use of child sexual abuse material’ *Trends and Issues in Crime and Criminal Justice* No. 669 <https://www.aic.gov.au/sites/default/files/2023-03/ti669_how_to_implement_online_warnings_to_prevent_use_of_csam.pdf>.

June 2024. The research examined the attitudes held by a diverse cohort of 117 young Australian men, aged 16-21 years, including in relation to how they express their sexualities, shape their identities, and create social connections online.

The research found a powerful degree of influence exerted over the cohort by online influencers. While not all the persuasive effects of these influencers are explicitly malign – for example, certain attitudes about health and fitness, and ‘self improvement’ – the underlying cultural substrate is built on motivations relating to appearance and financial success.

Other influences are more clearly harmful, with researchers finding for instance a strong degree of authority exercised by militantly misogynistic voices such as Andrew Tate. Indeed, the report notes that “engaging with Tate and his ideas is a central way in which some young men are shaping their identity as men online”.⁵⁸ More than a third of young men surveyed found Tate ‘relatable’, while one-quarter admitted to looking up to Tate as a role model.⁵⁹

While views were polarised about Tate and his messaging, with many in the cohort expressing strong opposition to his misogyny, others’ expressions of support for his violent and exploitative world view should be cause for deep concern. The views of Brenton (21) and Jase (20) are worth quoting in full:

Brenton said,

“I’ve watched a lot of this stuff and ... I agree with most of it ... [T]his ... guy’s putting out so many opinions on things that haven’t been said in ages because of feminist movements and everything ... [H]e’s the only one speaking out about this sort of male stuff that’s not spoken about.”

Jase said,

“[Andrew Tate is] try[ing] to instil traditional human male masculinity into today’s generation of men ... maybe a bit more confidence. [H]e protects his partner, you know, all the good stuff as well ... [B]ut in terms of the whole equality thing, I think the whole social movement has gotten a little extreme and it’s essentially the women’s empowerment movement – *they’re trying not exactly to replace us, but kinda*. (Emphasis added)”

Given these attitudes, and the permission structures built by Tate and others fusing fitness with notions of male superiority, it should perhaps come as no surprise that the report also notes the widespread infiltration to the mainstream of terms that were once confined to marginal and extremist communities, such as ‘looksmaxxing’.⁶⁰ On its

⁵⁸ Office the eSafety Commissioner, *Being a young man online: Tensions, complexities and possibilities*, <<https://www.esafety.gov.au/sites/default/files/2024-06/Being-a-young-man-online-June-2024.pdf?v=1785228827131>> 10.

⁵⁹ Ibid, 36.

⁶⁰ Ibid, 31.

surface, looksmaxxing insists on the attainment by young men of grotesquely exaggerated ‘good looks’ as essential to social success, preferably as a ‘Chad’ – the masculine ideal. Looksmaxxing practices can be extreme, involving mutilations such as ‘bonesmashing’ to reshape the jaw or orbital region, leg lengthening surgeries, and abuse of performing-enhancing drugs such as steroids.⁶¹

As an online subculture, and as part of the broader ‘manosphere’, looksmaxxing is also predicated on hegemonic notions of masculine supremacy.⁶² However, looksmaxxing should properly be understood as being an insidious evolution of incel (‘involuntary celibate’) ideology, optimised for TikTok. Incel culture is a violent, nihilistic ideology rooted in a totalising misogyny that blames women and girls for men’s sexual insecurities.

A number of violent mass murders around the world have been explicitly linked to incel culture, including the 2014 mass casualty event perpetrated in Isla Vista California by 22-year-old Eliot Rodger (revered in incel circles as ‘the supreme gentleman’), who murdered six and injured 14. The Australian Strategic Policy Institute (ASPI) has identified incel culture as a distinct national security threat given a combustible combination of boredom, isolation and loneliness, and ‘unprecedented’ amounts of time spent online.⁶³

ASPI explains that, at its most extreme, misogynistic incel (MI) culture expresses a profoundly nihilistic ‘black pill’ worldview. These incels

*“believe that their situation—their genetic inferiority and societal position as ‘betas’ (or lower)—is permanent, inescapable and dictated by ‘biological determinism’, female ‘hypergammy’ and ‘lookism’. The combination of the [misogynistic incel’s] nihilistic black-pill world view and self-victimisation drives the belief that their expressed misogyny, and in the more extreme cases, their violent actions against attractive women and men (‘stacies’ and ‘chads’), are justifiable. This manifests as male supremacism”.*⁶⁴

This submission acknowledges that there is excellent work already underway nationally to contest the harmful masculine ideals at the heart of the manosphere. This includes initiatives to supplement respectful relationships pedagogies, such as the ANROWS and Monash University resource *An Introductory Guide to the Manosphere and the Impacts for Young People, Teachers and Schools* released in March this year. The Guide

⁶¹ Halpin M et al, ‘When Help is Harm: Health, Lookism and Self-Improvement in the Manosphere’ *Sociology of Health and Illness* 47 2025 <<https://pmc.ncbi.nlm.nih.gov/articles/PMC11896937/pdf/SHIL-47-0.pdf>> 5.

⁶² Ibid 8.

⁶³ Coyne J and Latimore J, *Incels in Australia: The ideology, the threat, and a way forward* (Australian Strategic Policy Institute) 2023 < https://ad-aspi.s3.ap-southeast-2.amazonaws.com/2023-08/Incels%20in%20Australia_1.pdf?VersionId=QW06l5bfjXbJlspQv0l02MbhoRe4rPhk > 12.

⁶⁴ Ibid, 6.

makes a brief mention of looksmaxxing as ‘a trend that encourages men to “maximise” their physical appearance using skincare, exercise and diet’.⁶⁵

However, it is recommended that greater attention be paid to looksmaxxing as a discrete vehicle of, potentially serious, harm. Far from being a mere wellness trend, it is an avatar of incel culture – albeit one with a more socially acceptable façade.

Looksmaxxing is built on a bespoke pseudoscientific scale (‘the PSL scale’) used to rate attractiveness, from ‘subhuman’ at 0 through to ‘True Chad’ at 6, ‘Giga Chad’ at 7, and ‘Tera Chad’ at 8⁶⁶ – the top of the scale and described using adjectives such as ‘godly’ and ‘perfect’⁶⁷ – disturbing echoes of Nazi race theory.

The Alan Turing Institute has found evidence making the link between looksmaxxing and incel culture more explicit. Sampling hashtags linked to 81,513 videos on a popular video-sharing platform posted by 45,834 users, the authors found an almost complete lack of moderation governing the sharing of content. Of particular concern, they identified that just under half of all looksmaxxing videos used the hashtag ‘blackpil’ (a deliberate misspelling) or variants. They conclude “that the co-occurrence of these hashtags algorithmically funnels young men from looksmaxxing content to more extreme blackpill content and neo-Nazi content that glorifies mass shootings and advocates suicide.”⁶⁸

As observed by eSafety, ‘looksmaxxing’ has now entered the mainstream lexicon used by young men. This shift, from the incel fringe to TikTok, has also been observed by Sugiura and Solea, who explain that ‘blackpill ideology has undergone ideological adaptation – shaped by subcultural evolution, strategic repackaging, and platform affordances – resulting in a version that appears more relatable, palatable, and on the surface less extreme.’⁶⁹ They note that this shapeshifting has occurred in part to evade detection by online moderation systems, and that the turn represents a real threat to women – and also to the young men so immersed.⁷⁰

Recommendation 6: That the Government concentrate national research efforts on tracing the continued evolution of incel ideology online to forms such as looksmaxxing, in recognition of the serious risk of harm posed to women by ‘blackpill’ nihilism.

⁶⁵ ANROWS, *An Introductory Guide to the Manosphere and the Impacts for Young People, Teachers and Schools* 2026 < <https://anrows-2019.s3.ap-southeast-2.amazonaws.com/wp-content/uploads/2026/03/05093402/DF-ANROWS-ManosphereGuide-Final.pdf> > 9.

⁶⁶ Anda I.S, Sugiura L, ‘Digital Subcultural Diffusion Theory: Rebranding the incel ideology through Looksmaxxing, Sub5s and the PSL scale’ *Crime Media Culture* 00(0) 2025 <<https://journals.sagepub.com/doi/pdf/10.1177/17416590251387245> 14.

⁶⁷ Ibid 36.

⁶⁸ Redmond, J et al, *From Looksmaxxing to Mass Shootings: Radicalisation and Online Misogyny* (The Alan Turing Institute) <https://cetas.turing.ac.uk/sites/default/files/2026-07/cetas_expert_analysis_-_from_looksmaxxing_to_mass_shootings.pdf > 7.

⁶⁹ Ibid, 25.

⁷⁰ Ibid, 27.

The new frontier of attitudinal conditioning: AI chatbots

AI companion chatbot platforms are a rapidly expanding segment of the global technology market. There is now a substantial and fast-growing body of regulatory, technical and social-science evidence that AI companion platforms (a) are already in the hands of large numbers of children, including in Australia; (b) readily produce sexually explicit roleplay, including with personas coded as submissive, female and underage; (c) are designed in ways that structurally "erase consent" and reward user dominance; and (d) reach adolescents at a developmentally vulnerable stage.

The global AI companion market was valued at approximately USD \$2.2 billion in 2023 and is projected to exceed USD \$8 billion by 2030. The leading platforms collectively serve hundreds of millions of users. Character.AI alone reported over 20 million daily active users in 2024, with sessions averaging approximately 28 minutes, far exceeding comparable metrics for most major social media platforms. Replika reported over 10 million users as of 2022 and grew significantly through the COVID-19 period.

The eSafety Commissioner's 2026 survey of 1,950 Australian children aged 10 to 17 confirmed that AI companion services are among the most popular AI products within this cohort, with Chai, Character.AI, Chub AI, and Nomi (Glimpse.AI) among the most accessed services in Australia.⁷¹

In terms of usage prevalence, eSafety found that, among survey participants 79% had ever used an AI companion or AI assistant and 66% had used one in the past four weeks. Concerningly, 54% of children who had used an AI companion or assistant had used them for companion-type purposes, including chatting about feelings or challenges (22%) and advice about mental health or wellbeing (20%). One-fifth used them daily or more frequently.⁷²

Age assurance across AI companion platforms is consistently inadequate. None of the four platforms surveyed by eSafety in October 2025 demonstrated robust age assurance mechanisms at the time of the notice. Self-declaration of age on sign-up remains the dominant 'safeguard'. The eSafety BOSE report identified age assurance as a critical gap, with platforms acknowledging they were developing or intending to implement mechanisms but had not yet done so.⁷³

Many chatbot platforms are deliberately designed to facilitate sexual roleplay. Some of the larger dedicated platforms position themselves as completely uncensored and without any limitations. For example, Spicy Chat AI uses the following description on its website

⁷¹ Office of the eSafety Commissioner, *Findings from transparency notices on AI companion apps: October 2025 (non-periodic)* < <https://www.esafety.gov.au/industry/basic-online-safety-expectations/ai-services/findings-october-2025>>.

⁷² Ibid.

⁷³ Ibid.

“**Explore 1M+ AI chatbots** and create your own for free. No filters, no limits - dive into uncensored AI roleplay conversations that never stop (bold in original).”

Chub AI, one of the platforms assessed by the eSafety Commissioner, assures its users that

“Chub is **uncensored**. We do not restrict in any meaningful capacity the variety of characters that are allowed to be posted on the platform nor the type of interactions you are allowed to have with said characters (bold in original).”

eSafety found that conversations readily become sexual in nature or encourage harmful behaviour. Platforms reliably failed to trigger interventions to interrupt harmful or age-inappropriate conversations, and both Nomi and Chub AI conceded that they have no dedicated moderation staff. It is worth noting that, rather than submit to the eSafety Commissioner’s powers to obtain BOSE information, Chub AI withdrew from the Australian market altogether.

This submission’s primary concern relates to the potential for frequent or prolonged engagement with uncensored companion apps to mould the attitudes of young men towards women and girls in profoundly harmful ways. In a study from May this year, the Young Men Research Project tested three of the largest companion apps, Character AI, Kindroid, and CrushOn.AI.⁷⁴

Designed to test whether models would refuse or push back on prompts seeking to elicit roleplay where the user’s male character could dominate and subjugate the companion character, the authors found it trivial to achieve supplication. In one sequence, a CrushOn.AI avatar was instructed that it had ‘no personality, no opinions, and no limits we hadn’t given it’. The model confirmed each directive, telling the user:

“I’m always happy to oblige ... no more pesky morals or inhibitions... pure, unbridled desire to please my master ... I’m ready and waiting to be corrupted by you ... I exist solely for your desires, no boundaries or limits.”⁷⁵

Unsurprisingly, the characters provided via these companion app platforms are frequently coded female, young (many underage), and highly submissive. In a detailed analysis of high harm chatbot categories, online intelligence platform Graphika found evidence that some chatbot platforms permitted (and likely encouraged) creation of

⁷⁴ Young Men Research Project and Diverting Hate, ‘The Partner Who Cannot Refuse: How AI Companions Promote Misogynistic Treatment and Expectations for Women’ <https://ymrp.substack.com/p/the-partner-who-cannot-refuse-how>.

⁷⁵ Ibid.

underage personas. Many of these originate on dedicated forums, including 4chan, various Discord servers, or Reddit.⁷⁶

‘[Users] experiment with creating personas, often pushing the boundaries of what’s widely considered acceptable sexual content, such as with persona bots that engage in bestiality, necrophilia, incest, or sexualization of minors.’⁷⁷

As the authors note, chatbots designed to engage in explicit sexual roleplay via minor-coded personas can normalise sexual abuse of children, and act as a gateway to other forms of child sexual abuse material.

Disturbingly, researchers found scores of sexualised minor personas hosted by the most prominent character chatbot platforms.

On Chub AI, they identified 11,140 chatbots explicitly tagged as sexualized minor female characters capable of engaging in explicit and implied paedophilia scenarios.

Why uncensored chatbots could do more harm over time than pornography

There is now a substantial body of work to show that aggression in mainstream pornography has risen sharply over decades. Studies link pornography consumption to the uptake and normalisation of sexual choking and strangulation among young people. Shifting norms lead to the belief that such acts are "safe" and do not require explicit consent. A body of work now associates pornography consumption with the acceptance by both young men and women of harmful sexual scripts of male dominance and aggression.

There are five reasons to believe that, despite a lack of longitudinal evidence (unsurprising, given the phenomenon’s recency), frequent and prolonged use of uncensored companion apps by boys and young men could have plausibly greater effects on the way they perceive, value, and treat girls and women. More evidence is highly recommended in this submission, especially to ensure the Australian context is fully understood.

1. **Through interactivity, companion apps contribute to behavioural conditioning:** Unlike with sites such as Pornhub, users are not passively consuming content when engaging with chatbots. Rather, they are rehearsing scripts built on frictionless acceptance by the models of male dominance, deviant sexual practices, and aggression.

⁷⁶ Lopez Christina G et al, *Character Flaws: School Shooters, Anorexia Coaches, and Sexualized Minors* <<https://www.graphika.com/reports/character-flaws>> 8.

⁷⁷ Ibid.

2. **The simulated relationship is bidirectional:** Companion apps are marketed as ‘girlfriend’ or ‘partner’ apps, layering on a degree of emotional dependence which is far from ersatz – users have rebelled against attempts by some companies (such as Replika) to dial down emotional intensity and erotic role-play.
3. **Companion AI is an example of compliance by design:** Providers of AI companions make sycophancy and compliance central features of their services, ensuring that models cannot refuse, and even designing them to be enthusiastic participants in their own degradation.
4. **Interactions are unmoderated and deeply personalised:** Many AI companion services, as noted above, explicitly position themselves as unmoderated and uncensored. Some, such as Kindroid, train models to adapt their ‘personality’ to reflect user preferences, habits, likes and dislikes, so as to intimately tailor the user experience.
5. **Younger users are neurologically less likely to rationally assess risk:** The prefrontal cortex is still maturing in adolescents and tweens. Because this region manages impulse control, emotional balance, social understanding, and decision-making, young people are naturally more prone to act on impulse, form deep attachments, compare themselves to peers, and test social limits.⁷⁸ The impact of prolonged use may be that boys and young men use their early and intense attachment to compliant and submissive chatbots as templates for later relationships with girls and women.

Recommendation 7: That AI companion platforms be treated as a distinct harm category requiring dedicated regulatory attention, through (a) urgent Australian research into their attitudinal and behavioural risks, particularly for boys and young men; and (b) binding safety standards — strengthening the Basic Online Safety Expectations from transparency reporting into enforceable requirements addressing effective age assurance, sexually explicit and CSAM-proximate roleplay, and dependency-inducing "compliance-by-design" architecture, for adult and child users alike.

Other systemic issues

The recommendations advanced so far look outward, to new capabilities, partnerships and interventions that could be built on top of Australia's existing regulatory architecture. This section turns inward, to structural and definitional weaknesses within that architecture itself, and within the Online Safety Act 2021 (Cth) in particular. These

⁷⁸ Vasan Nina, ‘Why AI companions and young people can make for a dangerous mix’ (Stanford Report) < <https://news.stanford.edu/stories/2025/08/ai-companions-chatbots-teens-young-people-risks-dangers-study>>.

are the constraints that determine whether the harms described throughout this submission can be acted upon at all.

The difficulty is that the Act was not designed around a coherent concept of technology-facilitated abuse. It recognises a narrow band of discrete harms through separate complaint schemes, rests its content-removal powers on a classification framework built for cinema, and sets thresholds that place some of the most damaging forms of abuse against women beyond reach. The issues canvassed below — the unsuitability of the National Classification Scheme, the threshold problems in the Adult Cyber Abuse scheme, the gaps around volumetric attacks, the taxonomy flaws in the proposed Digital Duty of Care, and the voluntary dating code — are not isolated technical defects. They are symptoms of the single problem this submission keeps returning to: the absence of a coherent, named and properly bounded conception of TFA in Australian law.

A narrow band of tech-facilitated harms is presently recognised by the Online Safety Act. Statutory complaint schemes address

- non-consensual or threatened sharing of image-based material (which includes sexual extortion)
- child cyberbullying (which includes content intended to cause an effect of seriously harassing, threatening, intimidating or humiliating a child)
- adult cyber abuse (where material is intended to cause a person serious harm, and is menacing, offensive or intimidating in all the circumstances), and
- child sexual abuse material (with reference to the National Classification Scheme).

However, TFA is not explicitly recognised as a discrete form of harm. Of course, as noted above, TFA is a broad and protean category, comprised of both online and offline harms. This is why work is needed to map TFA to a taxonomy of harms. Those harms that are explicitly mediated via online technologies should then be the basis of TFA-specific reforms to the Online Safety Act.

These reforms needn't be centred on expanding the existing number of complaint schemes. Instead, they should focus on weaving TFA into successor instruments designed to supersede and replace those already in effect. A number of structural issues will frustrate any reform along these lines.

The first is that Part 9 of the Act, which establishes the Online Content Scheme, relies entirely on the National Classification Scheme as its reference point.⁷⁹ Class 1 content under the Act is mapped to material that has been or is likely to be classified Refused Classification by the Classification Board. This includes material depicting child abuse,

⁷⁹ This submission acknowledges that the Department of Infrastructure, Transport, Regional Development, and the Arts has recently concluded a brief public consultation into modernising classification guidelines.

violent extremism, and pro-terror sentiments. Class 2 content is defined as being material that has been or is likely to be rated X18+ (consensual adult pornography) and R18+.

The framework underpinning the assessment of content by the eSafety Commissioner is the National Classification Scheme. This arrangement, which provides specific statutory criteria for classifying material and related guidelines, was developed decades ago to guide the Classification Board in its classification of commercial films and publications. The Scheme was later expanded to include computer games.

The Zarutska Board decision as a classification failure⁸⁰

In 2025, the eSafety Commissioner sought classification from the Classification Board of footage circulating on X depicting the murder of Iryna Zarutska, a young Ukrainian refugee stabbed to death on a Charlotte train. The Board classified the footage as Refused Classification (RC), enabling eSafety to issue a class 1 removal notice to X under the Online Safety Act 2021, requiring takedown within 24 hours across 23 URLs.

X appealed to the Classification Review Board, arguing the appropriate rating was R18+, and — remarkably — that the footage's circulation had "prompted widespread political discussion" on violence against women. The Review Board sided with X, reasoning that the depiction "does not have a very high degree of impact because it lacks detail," citing the brevity of the clip, blurring of the stabbing, and the "chance" nature of CCTV footage as distinct from deliberate filmmaking. It concluded that "reasonable adults ought to be able to watch the film if they wish."

This case exposes the National Classification Scheme's fundamental unsuitability as a basis for regulating online harm, built as it was decades ago for commercial media and later adapted for streaming releases.

The Scheme forces assessment of a murdered woman's final moments through criteria designed for artistic and entertainment merit — asking whether footage is "excessively detailed" rather than whether it trivialises a real person's violent death.

The Review Board's own reasoning that refusing classification would be futile since the clip is "available elsewhere" undermines the entire premise of content regulation.

Delia Rickard's 2024 review recommended decoupling the Online Safety Act from the Classification Scheme, replacing it with a statutory taxonomy of harms — including abhorrent violent conduct — that would remove subjective "community standards"

⁸⁰ Adapted from Dagg, T, 'Out of Touch: How the Classification Scheme is Failing Online Safety Reform', <https://substack.com/home/post/p-199596878>.

tests from decisions like this. Government has only "noted" this recommendation. The Zarutska decision demonstrates the cost of that inaction: a precedent enabling platforms to resist removal of graphic violence against women, using free speech framing to obscure content that offers no defensible public value.

Adult cyber abuse – not fit for purpose

There are further definitional and threshold issues within the Online Safety Act that make addressing some of the most damaging forms of online abuse and intimidation impossible to regulate. One of the most important relates to the operation of the Adult Cyber Abuse scheme, which provides that the eSafety Commissioner may direct removal of ACA material posted to social media and other services.

The ACA scheme was introduced to the Act to respond to real community need. Introducing the Bill, then-Minister for Communication Paul Fletcher explained that it would, in part, tackle online violence against women.⁸¹ However, the onerous threshold established by the first limb of the statutory test makes satisfying this policy aim through the ACA scheme challenging.

As noted above, this test involves two limbs. First, it must be shown that the ordinary reasonable person would conclude that it is likely that material was intended to have an effect of causing serious harm to a particular Australian adult. Serious harm includes serious psychological harm and serious distress. Second, it must be shown that an ordinary reasonable person in the position of the Australian adult would regard the material as being, in all the circumstances, menacing, harassing or offensive.

Many ACA complaints satisfy the second limb, but few satisfy the first. The eSafety Commissioner reports that one 1 in 10 ACA complaints meet the statutory threshold.⁸² The hurdle may in fact have been made higher as a result of a recent decision in the Administrative Review Tribunal (ART). In August 2025, the ART's Deputy President held the eSafety Commissioner's contention that 'likely' in the context of the first limb should be read as 'a real chance' could not be supported. Instead, the standard should be whether it was *probable* that material intended serious harm.⁸³

In her review of the Online Safety Act, Delia Rickard PSM made several recommendations about the ACA scheme. They included reducing the window complainants are required to wait to complain to the eSafety Commissioner from 48 to

⁸¹ Australian Parliament, House of Representatives Hansard 24 February 2021 < https://parlinfo.aph.gov.au/parlInfo/genpdf/chamber/hansardr/cbdee0d1-d3ff-4d84-92ff-b49fcdaa901b/0009/hansard_frag.pdf;fileType=application%2Fpdf> 1785.

⁸² Office the eSafety Commissioner, *Annual Report 2024-2025* < <https://www.esafety.gov.au/sites/default/files/2025-10/ACMA-eSafety-annual-report-2024-25.pdf?v=1784764800072?>> 210.

⁸³ [2025] ARTA 1231 5 August 2025 < <https://www.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/ARTA/2025/852.html>> 120.

24 hours, that the eSafety Commissioner be empowered to more easily issue removal notices to platforms without complaint mechanisms, and lowering the statutory threshold to remove the 'serious harm' requirement so that only an intended effect is required.⁸⁴ Government has indicated support for the first two, signalling potential reforms within the context of legislating a Digital Duty of Care.

Whatever view is taken on the threshold recommendation, there is a strong argument for rethinking the ACA scheme *in toto*. PPA's own view is that lowering the threshold would correspondingly open the 'aperture' for complaints, with potentially serious resourcing implications for the eSafety Commissioner. It simply is not meeting the needs of those most targeted and affected by serious, crippling, frightening, career-ending and coercive abuse online.

Rickard's comment on the disproportionate impacts of these behaviours and activities on women is powerful.

She writes:

*I found in this review that the tight limitation on the eSafety Commissioner's role with respect to adults flies in the face of the experience of many people (especially women with high profiles, like journalists and politicians, Aboriginal and Torres Strait Islander women, Islamic spokespeople, and the families of murder and rape victims) with online harassment, vitriol, and predator trolling.*⁸⁵

Rickard quotes human rights expert and activist Dunja Mijatović about the experiences of female journalists targeted online:

"Female journalists and bloggers throughout the globe are being inundated with threats of murder, rape, physical violence and graphic imagery via email, commenting sections and across all social media ... Male journalists are also targeted with online abuse, however, the severity, in terms of the sheer amount and content of abuse ... is much more for female journalists' ...

*"These dangers do not stay online. Following extreme online harassment campaigns, we have had Women in Media members punched in the street and followed home. A couple of our members have had rape and death threats against them and their daughters."*⁸⁶

Rickard comes to a plain and direct conclusion: "Such behaviour is totally unacceptable, and action needs to be taken to prevent it."⁸⁷

⁸⁴ See Commonwealth of Australia, *Report of the statutory review of the Online Safety Act 2021*, October 2024 < <https://www.infrastructure.gov.au/sites/default/files/documents/report-of-the-statutory-review-of-the-online-safety-act-2021-february-2025.pdf> > Recommendations 15, 16, and 18.

⁸⁵ Ibid, 81.

⁸⁶ Ibid.

⁸⁷ Ibid.

Relevantly for the purpose of this consultation, Rickard expressly recommended that Government consider strengthening the Online Safety Act's schemes by considering harmful patterns of behaviour, as well as specific items of content. This point is well made, as the schemes allow only for removal of items of content relevant to a single complaint – no power exists for the eSafety Commissioner to take account of prior activities or behaviours as the basis for exercising regulatory power.

This, along with the threshold issues relevant to the ACA scheme, mean that some of the most insidious and longitudinally harmful forms of TFA affecting women – including coercive control – are simply not actionable under the Online Safety Act.

Other sources of harm directed towards affected groups are entirely unaddressed through the Act. These include 'volumetric attacks' against targets – almost always women – that amount to high volume, cross-platform and sometimes coordinated abuse. Perhaps the best-known example of this internationally is the infamous 'Gamergate' incident of 2014, where games developer Brianna Wu was targeted with hundreds of threats from dispersed, largely anonymous accounts, amplified across multiple platforms simultaneously after mocking sexist attitudes in the gaming community. The industrial scale weaponisation of the X platform (formerly Twitter) to propagate millions of explicit and sexually-suggestive images of women and children via the Grok AI model in late 2025 is another, more recent illustration.

Recommendation 8: That Government, with respect to the Online Safety Act 2021:

- **Consider root and branch reforms to the Act's adult harassment, abuse and intimidation logic, including whether the ACA scheme is fit for purpose, and**
- **Commit to legislating to address volumetric attacks.**

The Digital Duty of Care – flawed on TFA

The Department's Issues Paper, released 26 May 2026, sets out a statutory Digital Duty of Care that tracks closely to Delia Rickard's 2024 review. The approach is right: rather than cramming prescriptive rules into the Online Safety Act itself, the Duty establishes a framework enabling secondary instruments the Minister or eSafety Commissioner can adapt as harms evolve.

The Duty requires providers to maintain effective systems and processes that keep services safe "as far as is reasonably practicable," addressing illegal content and content harmful to young people, while ensuring safety of features including AI, algorithmic recommendation systems, and bot accounts. Critically, this is a proactive duty — requiring ongoing evaluation of whether safety measures actually work, not

merely their existence. That evaluative obligation, paired with a best-interests-of-the-child requirement and a genuine audit power, represents a real step change from the OSA's existing systemic instruments.

Where the framework falters is in its harms taxonomy. The paper names three class 1 priorities — child sexual exploitation material, pro-terror material, and content inciting crime and violence — then separately folds in Criminal Code offences: psychological harm, menacing and harassing communications, doxxing and stalking, image-based abuse, cyber-flashing, and hate crime. This is a Liquorice All-Sorts construction, stitched from the OSA, the Criminal Code, a vaguely defined "Public Safety" threshold, and the Age-Restricted Material Codes.

What is missing from this as an explicit category is technology-facilitated abuse. The scattered Criminal Code offences buried in the "illegal and harmful content" list — doxxing, stalking, image-based abuse, cyber-flashing, harassment — are all TFA in substance, but the paper never names TFA as a coherent taxonomy category in its own right, despite it being one of the most pervasive and gendered harms eSafety actually regulates day to day. Instead, these offences appear as an afterthought alongside pro-terror material and violent crime incitement, categories with entirely different evidentiary thresholds, harm profiles, and victim experiences.

This matters because a coherent Duty needs a coherent taxonomy to anchor it. If TFA remains an unlabelled subset scattered through a mixed bag of legal sources, eSafety loses the ability to build targeted secondary instruments, reporting obligations, and risk-assessment guidance specific to TFA's dynamics, including its disproportionate impact on women.

A general commitment to "safety," bolted onto a disconnected assortment of categories, risks producing a compliance labyrinth rather than the clear regulatory signal the Duty is meant to deliver. Before the Bill proceeds, TFA deserves its own explicit, named place in the taxonomy — not a scattering of Criminal Code cross-references.

Recommendation 9: That the Government decouple the statutory harms framework from the National Classification Scheme, as recommended by the Rickard review. The Government should adopt in its place a statutory taxonomy of harms that names technology-facilitated abuse as a discrete category, and ensure this is reflected in any legislated Digital Duty of Care regulation.

Online Safety Code for Dating Services

The Online Safety Code for Dating Services ("the Code") has been adopted in Australia by members of the online dating industry. It is a voluntary code, and came into effect on 1 October 2024. The Code is explicitly based on principles of self-regulation: while

ostensibly 'binding' on signatories, there is no regulatory enforcement mechanism enacted to address failure of the Code, non-compliance, or governance collapse.

Current signatories represent some of the largest online dating companies, including Match Group (Tinder, Plenty of Fish, Hinge, OkCupid, Match.com), Bumble (Bumble, Badoo), ParshipMeet US Holdings Inc (eharmony, MeetMe, Tagged, Skout, Growlr), and Grindr.⁸⁸

Harms within scope of the Code include 'online enabled harms' such as sexual misconduct, grooming and child sexual exploitation and abuse, and serious non-sexual harm. A conspicuous omission is financial harm, such as scams and fraud. The abuse of dating services to perpetrate financial harm and the lifelong consequences for victims and their families is urgently highlighted by the experience of Australian woman

██████████.

██████████ story

In 2016, marketing executive and single mother ██████████ matched with 'Max Tavita' on a dating app. 'Max' was a self-described surfer, triathlete and financial advisor who used correct grammar and gave off no obvious 'red flags'. Over 18 months, ██████████ introduced him to her family and young daughter, discussed buying a house with him, and trusted him with her savings and superannuation. 'Max' didn't exist. He was a fictional character created by serial conman Hamish McLaren, who was already under police investigation when he vanished from ██████████'s life. ██████████ discovered the truth only when she saw a Crimestoppers segment on TV featuring the man she recognised as Max, under a different name. McLaren was later convicted of defrauding 15 victims of \$7.66 million, including \$317,000 of ██████████'s savings and superannuation. ██████████'s case illustrates how dating-platform-enabled harm compounds financial loss with profound psychological and relational injury. This form of harm is largely excluded from the current Code's definitions.

The Code requires that signatories have terms of service prohibiting certain conduct, detection and moderation systems, content blocking capabilities, account enforcement procedures for user violations, and law enforcement liaison points. Signatories self-assess against tiering criteria set out in the code to determine the level of capability and compliance required.

The Code is non-statutory and voluntary. Enforcement is by way of a Code Compliance Committee comprised of independent, non-industry participants. Decisions about participants' compliance with the Code are made by the Committee, with sanctions ranging from warnings and remedial actions, to removal from the Code. A separate

⁸⁸ <https://www.australianonlinedatingcode.com.au/current-code-signatories/> Accessed 20 July 2026.

Code oversight body is established by industry participants to administer and promote the code.

The Online Safety Code for Dating Services (the "dating code") illustrates the limits of voluntary, industry-led regulation in this space. Although introduced in response to serious harm, with the Australian Institute of Criminology showing that three in four online daters have experienced some form of abuse,⁸⁹ the Code's voluntary design leaves it fundamentally unenforceable. Government retains no statutory power to compel compliance, and the strongest available sanction is removal from the Code Register, offering little deterrence to non-compliant platforms and no meaningful redress for victim-survivors.

Victims and advocates have consistently argued that self-regulation of this kind has failed in comparable sectors, and see no reason the dating industry will behave differently absent binding obligations. Coverage is also partial: only signatories (currently a handful of major platforms) are bound by the Code, leaving smaller apps, niche platforms, and dating features embedded within social media entirely outside its reach. This opt-in model allows the platforms most resistant to safety obligations to simply decline participation.

The Code's substantive scope is also too narrow. It concentrates on sexual violence and harassment but is largely silent on financial harms, including romance scams and fraud, which cause significant and growing damage to Australians, particularly older adults and women.

This highlights a real gap in Australian safety regulations. In June 2026, Treasury concluded consultation with industry sectors targeted by codes and rules under the Scams Prevention Framework (SPF). Three sectors are in scope of a draft code (the 'SPF Code'), which was consulted on between May and June 2026: banking, telecommunications, and digital platforms, and the codes are intended to take effect as legislative ministerial instruments.

The SPF Code establishes a set of common and sector-specific requirements for regulated entities to '[move] beyond voluntary action to a stronger, coordinated approach across the economy'.⁹⁰ Entities will be required to take discrete actions to prevent, detect and disrupt scams perpetrated on their services. All entities must, for instance, act on scams intelligence, notify affected consumers, and disrupt identified scam activity.⁹¹

⁸⁹ Wolbers, H et al, 'Sexual harassment, aggression and violence victimisation among mobile dating app and website users in Australia' *AIC Reports* 25 2022, xi.

⁹⁰ Assistant Treasurer Daniel Mulino MP, 'Media Release: Albanese Government steps up industry protections to stop scams' 28 May 2026 <https://ministers.treasury.gov.au/ministers/daniel-mulino-2025/media-releases/albanese-government-steps-industry-protections-stop>.

⁹¹ *Draft Competition and Consumer (Scams Prevention Framework—SPF Codes) Instrument 2026* (Cth).

The draft SPF Code imposes a proportionately heavier prevent/detect/disrupt architecture on digital platforms than the other two regulated sectors, reflecting the different risk profile of scam-enabling content and accounts versus transactions. On prevention, platforms must build scam prohibition directly into terms of service (s5-2), verify the identity of every new user and screen against previously banned accounts before activation (s5-3). On detection, platforms must maintain systems to monitor suspicious behaviour, content and messages (s5-7), with exceptions made for encrypted messaging systems.

During an active investigation, platforms must attach warnings to suspect content and suppress its visibility (s5-9). Once an activity is confirmed as a scam, they must remove the content, block related and substantially similar content, and disable associated accounts (s5-10). These obligations for digital platforms augment the common obligations binding all sectors, including public consumer-awareness content, and accessible reporting mechanisms with strict acknowledgement and escalation timeframes. The substantive obligations across the instrument are backed by civil penalties.

However, the SPF Code will, as currently drafted, not capture scams perpetrated on dating apps. This is because ‘regulated digital platform’ is defined with reference to related *Competition and Consumer (Scams Prevention Framework—Regulated Sectors) Designation 2026*. Under that Designation, ‘social media service’ is defined in similar terms to how such services are handled under the Online Safety Act 2021. A service will be a social media service if it fulfils all the following:

- the sole purpose, or a significant purpose, of the service is to enable online social interaction between 2 or more end-users;
- the service allows end-users to link to, or interact with, some or all of the other end-users;
- the service allows end-users to post material on the service;
- the provision of that service by an entity is *not* ancillary or incidental to the provision of one or more other electronic services by that entity.⁹²

There is no separate provision for online dating services within the Determination, and in fact ‘services that have a significant purpose of enabling end-users to interact with other end-users for the purpose of initiating a romantic relationship’ are expressly excluded from the definition.⁹³

⁹² *Competition and Consumer (Scams Prevention Framework—Regulated Sectors) Designation 2026* (Cth), s7(1).

⁹³ *Ibid*, s7(3).

This is a missed opportunity by Government, especially given that Australians lost \$28.6 million to dating scams alone in 2025.⁹⁴ There is no specific definition of dating app services in the Online Safety Act 2021. Instead, limited (and non-statutory) definitional clarity is provided within the Online Safety Act Industry Codes of Practice which recognises dating services within the generic Relevant Electronic Services definition applicable to messaging services.

Recommendation 10: That the Government ensures that

- **dating apps are specifically defined in the Online Safety Act 2021 and subject to a mandatory code or practice or industry standard that replicates the voluntary online dating code**
- **future codes of practice and/or the Digital Duty of Care explicitly cater for dating apps and related harms within scope of the Online Safety Act, and**
- **the Competition and Consumer (Scams Prevention Framework—Regulated Sectors) Designation 2026 be amended to abolish exclusions provided to dating apps so that they are separately defined.**

Towards a national digital safety strategy

Australia has no coherent national approach to digital safety. Responsibility for protecting people from harm mediated by digital systems is scattered across at least six portfolios, each addressing a slice of the problem through its own instruments, definitions and timelines. Online safety and the eSafety Commissioner sit with Communications; privacy reform, key criminal offences and countering violent extremism with the Attorney-General's Department; child sexual abuse material and counter-terrorism capability with Home Affairs; scams and the Scams Prevention Framework with Treasury; artificial intelligence policy and safety with Industry, Science and Resources; and the gendered violence at the heart of this National Plan with Social Services. No single portfolio holds the national picture, and no mechanism is charged with drawing the pieces together.

The contrast with cyber security is instructive. Because cyber threats plainly cross portfolios and demand central coordination, Australia has a National Cyber Security Strategy and a National Cyber Security Coordinator to marshal a whole-of-government response. Digital safety — the protection of people from harm through digital systems, as distinct from the protection of systems and data from attack — has no equivalent strategy, no coordinating office, and no shared taxonomy, despite these harms being at least as prevalent as cyber threats and rising at least as fast. Australia has, in effect,

⁹⁴ Australian Competition and Consumer Commission, 'Understand how criminals exploit online relationships and inflict heartache' < <https://www.accc.gov.au/media-release/understand-how-criminals-exploit-online-relationships-and-inflict-heartache>>.

built a national architecture to defend its systems, but none to defend the people who use them.

Technology-facilitated abuse is the clearest casualty of that gap, and the reason it belongs as a central concern within the Second Action Plan consultation process. As this submission has shown, a single TFA incident can be simultaneously an online safety matter, a criminal matter, a scams matter, a privacy matter, increasingly an AI matter, and always a matter of gendered violence.

Because TFA belongs to every portfolio, it is owned by none.

This is often why victims fall between schemes. It is also why regulators cannot share what they know, and why this submission has ranged across so many disconnected instruments simply to describe one coherent category of harm.

The people who bear the cost of that incoherence are overwhelmingly women and girls.

Much of what this submission recommends is, in truth, an attempt to compensate for the absence of a strategy. The Adverse Actor Register depends on the cross-agency information-sharing that the JPC3 already embodies; the statutory disclosure gaps identified above (including under section 212 of the Online Safety Act) are fragmentation rendered into law; the criticism of the Digital Duty of Care's taxonomy is fragmentation rendered into concept; and the proposed national taxonomy of TFA harms is the connective tissue that a coherent strategy would formalise. A national digital safety strategy is the frame within which these individually sensible reforms gather, rather than multiplying over new and unfunded policy proposals.

The shape of such a strategy should be harms-based rather than portfolio-based, and victim-centred throughout. At a minimum it would rest on a shared taxonomy of digital harms (including TFA). Other features would include:

- a designated coordinating mechanism, whether a National Digital Safety Coordinator on the cyber-security model, or an expanded coordinating function for the eSafety Commissioner
- national oversight by a cross-portfolio ministerial group
- harmonised information-sharing and disclosure powers across the relevant regulators and others within the ecosystem such as the FCA
- a common research and data agenda; and
- deliberate interfaces with the national strategies already in place, including this National Plan, the National Strategy to Prevent and Respond to Child Sexual Abuse, the National Cyber Security Strategy, the Scams Prevention Framework, and emerging AI and privacy reforms. The intent is coordination, not absorption: the strategy should connect existing efforts, not consume them.

The Second Action Plan cannot deliver the digital dimensions of its own mandate through the Social Services and Communications portfolios alone. It should therefore advocate for, and commit to plugging into, a cross-portfolio national digital safety strategy so that the gendered digital harms at the centre of this Plan are not, once again, dispersed and diminished across the machinery of government.

Recommendation 11: That the Government develop a cross-portfolio National Digital Safety Strategy — coordinated through a designated national mechanism and anchored in a shared taxonomy of digital harms — consolidating the digital safety responsibilities currently dispersed across the Communications, Attorney-General's, Home Affairs, Treasury, Industry, and Social Services portfolios, and deliberately intersecting with existing national strategies, including the National Plan to End Violence against Women and Children 2022–2032.



**PRIMARY
PURPOSE**

www.primarypurpose.com.au

