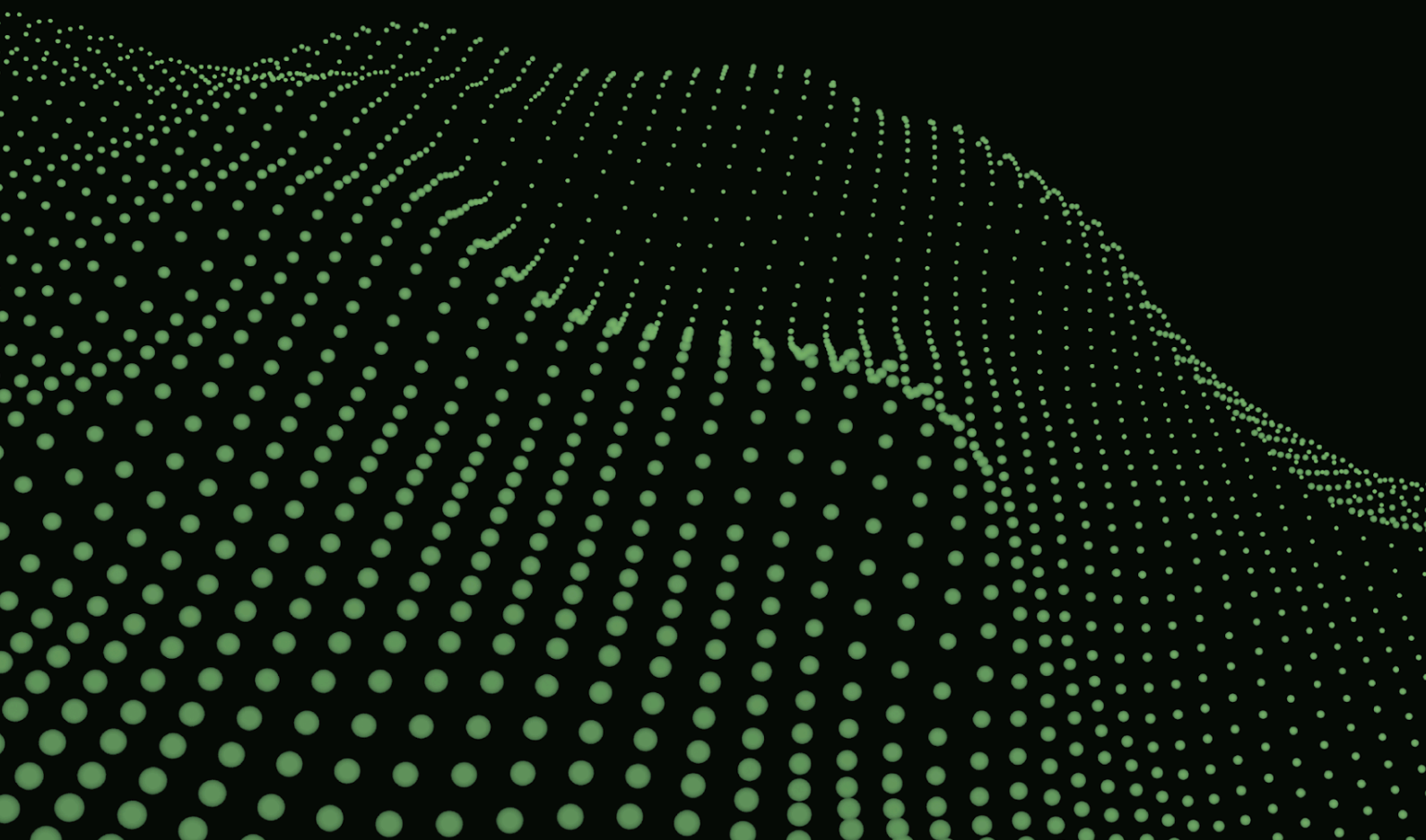
DIGI

Australian Code of Practice on Disinformation and Misinformation | **Annual Report**

Published June 10, 2026



Background

The *Digital Industry Group Inc (DIGI) is a non-profit industry association that advocates for the interests of the digital industry in Australia. DIGI's founding members are Apple, Discord, eBay, HelloFresh, Google, Meta, Microsoft, Pinterest, Snap, Spotify, TikTok, Twitch and Yahoo. DIGI's vision is a thriving Australian digitally-enabled economy that fosters innovation, a growing selection of digital products and services, and where online safety and privacy are protected. DIGI is a key Government partner in efforts to address online harms, data and consumer protection online and to grow the digital economy, through code development, partnerships and advocacy for effective and implementable approaches to technology policy.*

The Australian Code of Practice on Disinformation and Misinformation (ACPDM) was developed in response to Australian Government policy announced in December 2019 following the Australian Competition and Consumer Commission (ACCC) Digital Platforms Inquiry. The digital industry was asked by the Government to develop a voluntary code of practice on disinformation and news credibility signalling.

DIGI developed the ACPDM with assistance from the University of Technology Sydney (UTS) Centre for Media Transition, and First Draft, a global project that aimed to help societies overcome false and misleading information.

The ACPDM was launched in February 2021, and its current signatories are Apple, Google, Meta, Microsoft, Redbubble, TikTok, and Twitch. Adobe withdrew from the code in 2026. Adobe has informed DIGI that it remains committed to advancing the work it has led for more than six years with the Content Authenticity Initiative (CAI), C2PA steering committee and provenance solutions via Content Credentials.

The Code includes mandatory code commitments by all signatories to publish & implement policies on misinformation and disinformation, provide users with a way to report content against those policies and to implement a range of scalable measures that reduce its spread & visibility (Mandatory commitment #1). Every signatory must provide annual transparency reports about those efforts to improve understanding of both the management and scale of misinformation and disinformation in Australia (Mandatory commitment #7).

Additionally, the Code includes a series of opt-in commitments that platforms adopt if relevant to their business model: (Commitment #2) addressing disinformation in paid content; (#3) addressing fake bots and accounts; (#4) transparency about source of content in news and factual information (e.g. promotion of media literacy, partnerships with fact-checkers) and (#5) political advertising; and (#6) partnering with universities/researchers to improve understanding of misinformation and disinformation.

DIGI produces this annual report as part of its governance of the ACPDM.

Opening statement

By: Dr Jennifer Duxbury

Director, Policy, Regulatory Affairs and Research DIGI

2026 Annual Report Overview

It is my pleasure to present the 2026 DIGI Annual Report for the Australian Code of Practice on Disinformation and Misinformation (ACPDM) highlighting the ongoing commitment of the ACPDM signatories to fostering a more reliable and resilient digital environment for everyone in Australia. We have provided our most comprehensive overview to date of signatories' initiatives and their progress under the Code throughout the 2025 calendar year. Our aim is to help inform policymakers, industry partners, and the general public with insights and data that shed light on the ongoing efforts of signatories to combat disinformation and misinformation on their services.

Combating disinformation

Throughout 2025, ACPDM signatories made notable progress in identifying and mitigating disinformation risks with major platforms proactively disrupting Coordinated Inauthentic Behaviour (CIB) operations and publicly disclosing threat indicators. By providing transparency around tactics, origins, and emerging campaign patterns, large platforms can assist democratic governments protect themselves from disinformation threats globally. While Australia has not been subject to a large-scale foreign disinformation campaign to date, signatories remain alert to this risk, including the potential for bad actors to exploit Artificial Intelligence (AI) to spread false and misleading content. The 2025 Australian Federal Election provided a significant test of signatories' responses to disinformation and misinformation at a critical moment. Australia is one of the most online electorates in the world today. The October 2024 Interim Report from the Senate Select Committee on Adopting Artificial Intelligence underscored the committee members' concern that generative AI may be misused to rapidly spread false material aimed at undermining the integrity of democracy electoral processes and wider public discourse in Australia. During the 2025 campaign period, major platforms deployed targeted programmes and dedicated significant resources to support the work of the Australian Electoral Commission in ensuring the integrity of the electoral process.

Improving the digital information environment

Signatories continue to harness AI to improve the information environment on their services, deploying innovative tools to improve both content moderation and content authenticity including LLM-enabled comment moderation on MSN by Microsoft, Google's SynthID, and AI-labelling initiatives

by Meta and TikTok. As in 2024, an ongoing theme across this year's signatories' reports is a growing focus on the provenance and authenticity of digital content. Signatories' reports show how platforms can play a critical role in helping users understand the origin and nature of content they encounter online This work is complementary to signatories' systems and processes that proactively detect and reduce the spread of disinformation and misinformation material.

Collaboration and Research

Signatories have expanded their engagement with industry partnerships, academic research, and media literacy programs. This collaborative approach is increasingly recognised internationally as essential to understanding and responding to mis- and disinformation, including in the Organisation for Economic Co-operation and Development (OECD)'s December 2024 Recommendation on Information Integrity. DIGI's own evidence to the Select Committee on Information Integrity on Climate Change and Energy similarly underscored the value of multi-stakeholder cooperation in developing effective, proportionate responses.

Governance, Transparency, and Code Reform

The debate surrounding the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024 and its subsequent withdrawal have prompted reflection by DIGI and Signatories on how best to address misinformation and disinformation in ways that protect democratic discourse without restricting legitimate public debate. DIGI and ACPDM signatories conducted a comprehensive review of the Code, commencing in 2025, which was informed by signatories' feedback, the findings of the Select Committee on Information Integrity on Climate Change and Energy and broader stakeholder input. The review focused on strengthening the Code's key outcomes, governance frameworks and transparency arrangements. The revised Code and governance arrangements came into effect on 30 May 2026 and is discussed in more detail in this annual report.

2025 report' key insights:

- **Google:** Search's 'About This Result' feature was viewed more than 65 million times in Australia in 2025, providing users with transparency on why a specific web page appears in your search results and verifies the credibility of the website.
- **Meta:** Voter Information Units reached approximately 11.8 million Facebook users and 9.5 million Instagram users during the 2025 Federal Election, directing users to authoritative information about how, when and where to vote.
- **TikTok**: Maintained proactive detection and removal rates exceeding 97% across all four quarters for misinformation policy violations in Australia which it attributed to enhanced machine-based content recall and continued advancements in its operational and moderation capabilities.
- **Microsoft**: LinkedIn blocked more than 1.4 million fake accounts attributed to Australia.
- **Redbubble**: Achieved an 81% reduction in the reach of misinformation to Australian users on its platform.
- **Twitch:** Resolved all 31 Australian user reports related to misinformation within 24 hours.

The quality of the 2025 reports reflect signatories' commitment to accountability and continuous improvement. DIGI and ACPDM signatories look forward to working with government, civil society, researchers, and the public to further strengthen the Code's impact in 2026 and beyond.

Adobe

We note that Adobe has chosen to discontinue its participation in the Code. Adobe has informed DIGI that it will continue to focus its efforts on advancing content authenticity and provenance solutions globally through the Content Provenance Initiative, C2PA steering committee and provenance solutions via Content Credentials.

Table of contents

List of Acronyms

The following acronyms are used in this report and signatories' transparency reports.

Acronym	Definition
Organisations & Bodies	
ACCC	Australian Competition and Consumer Commission
ACMA	Australian Communications and Media Authority
AEC	Australian Electoral Commission
AAP	Australian Associated Press
AFP	Agence France-Presse
ASPI	Australian Strategic Policy Institute
DIGI	Digital Industry Group Inc
ECANZ	Electoral Council of Australia and New Zealand
GIFCT	Global Internet Forum to Counter Terrorism
IFCN	International Fact-Checking Network
NLP	News Literacy Project
OECD	Organisation for Economic Co-operation and Development
QUT	Queensland University of Technology
UNESCO	United Nations Educational, Scientific and Cultural Organization

Acronym	Definition
UTS	University of Technology Sydney
WSU	Western Sydney University
Codes & Programs	
ACPDM	Australian Code of Practice on Disinformation and Misinformation
C2PA	Coalition for Content Provenance and Authenticity
CAI	Content Authenticity Initiative
GPT	Generative Pre-trained Transformer (machine learning technology)
Technical & Policy Terms	
AI	Artificial Intelligence
AIGC	AI-Generated Content
CIB	Coordinated Inauthentic Behaviour
CIO	Covert Influence Operations
DSP	Demand-Side Platform
FIMI	Foreign Information Manipulation and Interference
GTIG	Google Threat Intelligence Group
OSINT	Open Source Intelligence
OSIT	Off-Service Investigation Team
RAG	Retrieval-Augmented Generation

Acronym	Definition
SIEP	Social Issues, Elections and Politics
Google-Specific Terms	
GTIG	Google Threat Intelligence Group
SynthID	Google's watermarking technology for AI-generated content
TAG	Threat Analysis Group
Microsoft -Specific Terms	
MTAC	Microsoft Threat Analysis Center
MSN	The Microsoft Network
MSTIC	Microsoft Security Threat Intelligence Center
Platform Features & Designations	
AIV	Advertiser Identity Verification
CCL	Content Classification Labels
GPPPAs	Government, Politician, and Political Party Accounts
TYA	Tune Your Algorithm

Part 1 | **Transparency reports**

Reports published on DIGI website for 2025 calendar year

The core objective of the annual transparency reporting process of the ACPDM is to provide the Australian public, the Australian Communications and Media Authority (ACMA), and the Australian Government with the means to evaluate signatories' adherence to their commitments under the ACPDM. The signatories' commitments to transparency under the Code are supplemented by Best Practice Reporting Guidelines that help signatories to more effectively articulate how they fulfil their obligations to combat disinformation and misinformation. Additionally, the transparency reporting process includes an independent review and assessment by Shaun Davies, who analyses signatories' draft reports, suggests improvements, and provides an attestation of the final published claims.

This year's transparency reports, available on the DIGI website, cover the period from January 1, 2025, to December 31, 2025. A summary of each report's insights can be found in Appendix A. Further contextual analysis of the signatories' actions to fulfil commitments under Outcomes 1 and 6 of the Code is provided below.

Key Themes from 2025 Transparency Reports

Artificial Intelligence: Expanding Use in Detection, Labelling and Content Integrity

Generative AI was the defining theme of the 2025 reporting period, with signatories expanding its use both as a tool to combat misinformation and as a subject of new policy and transparency obligations. The collective experience of Signatories, across the social media, search, professional networks, live streaming, marketplaces, and news services which are subject to the ACPDM, highlights widespread improvements in AI as a tool for enhancing platform integrity where, conversely, threat actors increasingly utilise AI to generate and disseminate disinformation.

Meta, TikTok, Google, and Microsoft significantly expanded AI-driven detection capabilities during the period. Meta reported that its detection systems remained robust against evolving Coordinated Inauthentic Behaviour (CIB) networks. Meta's enforcement strategies rely on highly accurate behavioural and technical signals to detect and remove CIB networks. Google reported that Gemini-powered tools enabled its systems to catch over 99% of policy-violating ads before they were served, and reduced incorrect advertiser suspensions by 80%. Across all four quarters, TikTok maintained proactive removal rates exceeding 97% for violations of its integrity and authenticity policies in Australia. Microsoft expanded AI-powered moderation on the Microsoft Network (MSN) and reported that MSN proactively blocked 988,892 policy violating comments in Australia through automated systems. Misinformation represented 0.34% of total comment takedowns by MSN, down from 0.77% in 2024.

Redbubble's experience illustrates how AI-assisted detection works in marketplace contexts. The platform's suite of scalable moderation tools, includes duplicate detection, image matching, keyword and text detection, and machine learning-based account abuse detection. Redbubble measures the reach of misinformation content by tracking "misinformation impressions".i.e. content users have engaged with before moderation. In 2025, deployment of moderation tools contributed to an 81% reduction in misinformation impressions among Australian users, from 49,496 in 2024 to 9,558 in 2025. The average time that violating content remained on the platform fell from 776 days to 139

days, suggesting faster detection cycles. Twitch, by contrast, did not observe generative AI materially changing the prevalence of harmful misinformation on its service in 2025, attributing this to the platform's community-based architecture, which inherently sees less amplification of AI-generated content.

All signatories made material advances on AI content labelling. Meta expanded its "AI Info" labelling across advertising surfaces. TikTok strengthened its labelling framework through creator disclosure tools, automated detection models, C2PA Content Credentials and invisible watermarking. Google deployed its SynthID watermarking technology across text, audio, image and video content, and its SynthID Detector portal for third-party verification. Microsoft and LinkedIn also adopted C2PA's "Content Credentials" open standard, and content containing the "Content Credentials" technology will be automatically labelled on LinkedIn, with users seeing the "Cr" label. Apple News requires that articles generated with AI assistance be labelled within the News app and encourages publishers to disclose how AI was used, with publishers subject to suspension for using AI to mislead readers. Many major platform signatories are members of the Coalition for Content Provenance and Authenticity (C2PA).

Election Integrity: Coordinated Platform Responses to the 2025 Australian Federal Election

The 2025 Australian Federal Election was a significant test of platform integrity measures, and Signatories' transparency reports illustrate a multi-layered response across their different types of digital services. The 2025 transparency reports show a range of approaches adopted by signatories with different business models. Larger platforms deployed large-scale operational programs, while others whose business models are less directly implicated by electoral advertising were focused on ensuring authoritative information was surfaced and harmful content was addressed.

Meta activated a cross-functional election integrity team, partnered with the Australian Electoral Commission (AEC), and deployed Voter Information Units that reached approximately 11.8 million Facebook users and 9.5 million Instagram users. Meta also removed over 65,000 ads in Australia for non-compliance with its social issues, elections and politics advertising policies during the reporting period. Similarly, TikTok mounted what it describes as its most extensive Australian election operation to date. Its Election Centre, developed in collaboration with the AEC and fact-checking partner AAP received over 400,000 views and its associated search guide viewed more than 1.9 million times. TikTok reported zero major content integrity escalations during the election period. Google, which has a focused election-content operation and allows only verified advertisers to run election ads in Australia, rejected 37,217 ads due to unverified advertisers.

Microsoft, with a large and diverse range of services, undertook heightened security monitoring of electoral systems through its MTAC, MSTIC, DCU and GHOST teams and also ran a public engagement session in January 2025 with over 150 people from Australia's political parties, newsrooms, academia and government. Additionally, Microsoft also enabled Bing to direct election-related queries to authoritative sources including the AEC.

The 2025 federal election was one of the most significant editorial undertakings of the year for Apple News Australia. Coverage under the "Election 2025" package was regularly featured in Top Stories throughout the campaign period, drawing from public broadcasters, major mastheads and digital outlets. On election night, Apple News delivered a real-time results experience built through a data partnership with The Australian, featuring two-party preferred estimates, an interactive seat-by-seat map, an electorate search function and a live seats-won tracker updated in near real-time alongside a

live feed of the ABC News channel. In the days following the election, editors continued updating the collection with outcome analysis, leadership responses and broader political commentary. Apple's approach, which prioritises accuracy and quality through deliberate human editorial curation, is a distinct and valuable contribution alongside the automated systems deployed by social media platforms.

Twitch, which prohibits political advertising in streaming and whose Harmful Misinformation Actor policy covers civic process interference, reported no significant election-related misinformation incidents during the period. Across all signatories, partnerships with the AEC and AAP FactCheck were central to election integrity operations, reflecting the value of pre-established relationships with authoritative domestic institutions.

Combating Coordinated Inauthentic Behaviour: Evolving Threats and Enforcement

The 2025 reports document the continued evolution in the tactics of coordinated inauthentic behaviour (CIB) networks, with signatories adapting their detection and enforcement approaches accordingly. The diversity of signatories' platforms highlights that the nature and scale of inauthentic behaviour varies significantly by platform architecture.

Meta removed 16 new covert influence operations globally during 2025, originating from China, Iran, Russia, Pakistan, India, Belarus, Moldova and Romania. Key trends identified by Meta included the proliferation of generative AI to create convincing personas and multilingual content at scale, and the development of two-tiered structures combining mass fake accounts with deeply developed "custom personas". Meta also observed CIB networks increasingly co-opting authentic voices and injecting narratives into the comment sections of prominent legitimate news pages.

TikTok reported no identified covert influence operation networks specifically targeting Australia in 2025, noting the platform continued to invest in expert teams using threat intelligence frameworks globally. Google's Threat Intelligence Group (GTIG) continued to monitor and disable accounts associated with coordinated influence operations, publishing quarterly updates.

LinkedIn took substantial action on inauthenticity globally, blocking more than 197 million fake accounts in 2025. A majority of fake accounts were stopped at registration. In Australia specifically, LinkedIn blocked more than 1.4 million fake accounts and removed 933 pieces of misinformation. LinkedIn's approach leverages advanced network algorithms, computer vision, natural language processing, and deep learning models to detect AI-generated profile elements such as deepfakes. Bing Search also deployed "defensive search" interventions in Australia across 324,965 queries and 6,358,885 impressions in April–December 2025, up from 225,062 queries in 2024.

Twitch's experience demonstrates that inauthentic behaviour manifests differently on live streaming platforms. In H2 2025, Twitch issued 42.7 million account enforcements globally for spam, scams, and fraud — 36,090 of which were for Australia-based accounts. The spike in Australian enforcement during H1 2025 (158,855 actions) reflected coordinated attempts to create large volumes of accounts for financial abuse, largely detected proactively by the platform at account creation. Most impersonation on Twitch took the form of phishing attempts via fraudulent channels, which Twitch addresses through URL scanning and proactive monitoring. Redbubble similarly focuses on account-level integrity. In 2025, over 495,000 accounts were disabled globally by its abuse detection software. This is a 65% increase compared to approximately 300,000 in 2024, including networks of connected accounts responsible for coordinated uploads of misinformation-related product listings.

A notable emerging disinformation threat documented in Meta's report was the potential for influence operations to engage in 'data poisoning' by designing content to shape AI training data toward specific narratives. Microsoft similarly identified the risk of its services being used to amplify foreign

cyber influence content and outlined six focus areas to combat the harmful use of deceptive AI. These developments signal a new frontier in information integrity risk warranting ongoing attention from all platforms at risk of this activity.

4. Research, Media Literacy and Multistakeholder Engagement

Signatories continued to support independent research and media literacy initiatives during 2025, with a notable concentration of activity ahead of the federal election and in response to the growing prevalence of generative AI. The range of approaches across signatories' services reflects how different platform contexts call for different models of media literacy engagement.

Meta partnered with AAP to deliver a multilingual media literacy campaign which was made available in English, Simplified Chinese, Vietnamese and Arabic. This initiative reached 823,638 Australians, accumulated over 5.07 million views and generated more than 2.45 million completed video plays. Meta also supported research from La Trobe University and the University of Waikato which drew on Meta's Content Library to analyse more than three million posts from 25 Australian news publishers to examine how news content is distributed, how audiences engage with news topics, and the way misinformation spreads. Meta also commissioned new research from Queensland University of Technology examining how Australians participate and engage with community-based content moderation systems (to be published in 2026). TikTok continued its partnership with AAP FactCheck, producing up to 29 debunking articles per month, and hosted two expert roundtables to map the election information landscape. Google supported a range of Australian initiatives, including as a founding supporter of Squiz Kids' Newshounds media literacy program, now active in over 2,000 Australian classrooms.

Microsoft contributed through several programs with direct Australian relevance, including a May 2025 pilot training for Australian content creators and influencers on AI ethics, the Reed Smart learning game on Minecraft Education, and its Search Progress information literacy tool adopted in more than ten countries. Microsoft also provided pro-bono advertising space across its surfaces to disseminate literacy campaigns, generating millions of impressions per month.

Twitch collaborated with media literacy expert MediaWise to develop educational materials for streamers and viewers on identifying and avoiding spreading misinformation online, hosted on the Twitch Safety Center. These materials received 1,886 global user visits during 2025. Twitch also promoted these resources through a Creator Camp livestream titled "Media Literacy with MediaWise," featured on Twitch's front page. This is a best practice example of a platform using its native creator ecosystem to amplify literacy messaging. Apple News' provides ongoing support for The News Literacy Project (NLP), a global organisation empowering young people with critical thinking skills for navigating digital news. This reflects Apple's commitment to news literacy which it views as a fundamental pillar of a healthy democracy, and to empowering young people to be informed citizens. Collectively, these initiatives reflect a multifaceted approach to supporting societal resilience to misinformation, spanning primary school curricula, adult media literacy campaigns, creator education, academic research access and high-level stakeholder dialogue

5. Transparency and Reporting: Progress and Challenges

All signatories published detailed transparency reports for the 2025 period, with improvements in the granularity and accessibility of Australia-specific data compared to prior years. TikTok added a dedicated Regional Measures page with an Australian focus to its Transparency Centre. Meta provided Australia-specific enforcement figures across multiple policy categories, noting that over 270,000 pieces of content were actioned for misinformation policy violations and over 4.8 million distinct pieces of content on Facebook received fact-check warning labels. Google continued to

publish quarterly YouTube enforcement data, quarterly GTIG bulletins and an annual Ads Safety Report with Australia-specific metrics.

Microsoft's 2025 report provided Australia-specific data across its services: 1.7 billion ad rejections in Australia, 77,307 total appeals and 722,299 entity takedowns through Microsoft Advertising; disaggregated defensive search intervention data for Bing in Australia for the first time; and year-on-year MSN misinformation takedown breakdowns by narrative category. LinkedIn continued to publish its Community Report twice yearly, and Microsoft also publishes its Digital Safety Content Report, Digital Defense Report, and Responsible AI Transparency Report through its Reports Hub.

Redbubble provided granular Australia-specific data, including quarterly moderation figures broken down by topic (conspiracy theories, medical misinformation, election misinformation), year-on-year impressions and clicks data, and user report volumes. Redbubble's reporting of reach metrics showing that harmful misinformation impressions fell 81% year-on-year to 9,558 in 2025. Redbubble's approach is a useful model for how smaller platforms can provide meaningful, outcome-focused transparency data.

Twitch published biannual safety transparency reports covering its Harmful Misinformation Actor Policy globally, as well as Australia-specific enforcement data for spam and inauthentic behaviour. In Australia during 2025, Twitch received 31 user reports related to misinformation concerns, all reviewed and resolved within 24 hours. Apple News provided global and Australia-disaggregated data on user-submitted concerns: of approximately 452,000 global concerns reported in 2025, around 3.5% originated in Australia, with approximately 52% of Australian reports categorised as misinformation or disinformation by the reporter. None of these concerns from Australia were substantiated by Apple's moderation team, reflecting the platform's curated, publisher-only model.

Notwithstanding this progress, DIGI acknowledges that a recurring challenge to transparency reporting is the limited comparability of data over time. Changes in platform policies can produce material discrepancies in figures e.g. around enforcement, that make it difficult to identify useful data points for year-on-year comparative analysis. TikTok similarly noted that the higher share of integrity violations in 2025 reflected a decline in overall enforcement volume rather than an increase in violating content. Microsoft also noted that year-on-year changes in advertising takedown figures reflect a shift toward proactive enforcement and methodology changes. Improving the consistency and comparability of transparency data remains an ongoing focus for the Code. In the Part 2 of this report we explain how following the 2025 code review, DIGI and signatories have sought to improve on the current reporting guidelines with recommendations on quantitative and qualitative datapoints.

Tailored Approaches Across Different Digital Environments

A distinguishing feature of the ACPDM is the diversity of the signatories. Alongside the large social media and search platforms that have historically anchored transparency reporting under the Code, the ACPDM signatories include a live streaming service (Twitch), a curated news service (Apple News), and an artist marketplace (Redbubble). Their reports collectively illustrate that effective approaches to combating mis- and disinformation are necessarily shaped by the nature of each platform's content, community and business model.

Twitch's approach is an example of how a smaller platform can deploy a targeted detection strategy with great success. The platform considers itself a lower-risk environment for misinformation, and its data supports this assessment. Global enforcement under its Harmful Misinformation Actor Policy declined from a peak of 90 in H1 2022 to zero indefinite suspensions in H2 2025. Twitch attributes this trajectory to the structural characteristics of its platform. Building an audience on Twitch is slow, content is long-form and ephemeral, and it does not readily lend itself to the rapid viral spread of false claims. Rather than deploying mass automated detection, Twitch focuses on persistent patterns of

behaviour evaluated across a streamer's on- and off-platform presence, supplemented by user reporting and a dedicated Off-Service Investigation Team (OSIT) for severe off-platform conduct.

Apple News' model is structurally distinct from other signatories: not being a user-generated content platform, Apple News' design choice is its primary contribution to combatting misinformation. Publishers are vetted by Apple's editorial team before onboarding, and the highest-visibility area of the app, Top Stories, is entirely curated by experienced Australian journalists who vet each article for accuracy, fairness and news value. In 2025, of the Apple News concerns identified as originating in Australia that were categorised by users as misinformation or disinformation, none were ultimately substantiated by the moderation team. Apple's coverage of both the 2025 federal election and the Bondi Beach terror attack in December 2025 demonstrated the platform's deliberate prioritisation of accuracy and quality during high-stakes news events.

Redbubble's experience highlights the specific challenges of marketplace platforms, where misinformation manifests not in posts or articles but in product listings for t-shirts, stickers, and prints carrying text-based slogans or imagery associated with conspiracy narratives. Conspiracy theory-related content, primarily QAnon and 9/11-related uploads, accounted for the majority of moderation activity in 2025. Redbubble's report documented a resurgence in QAnon content, fuelled in part by the return of Donald Trump to the US presidency and ongoing public interest in the Epstein files. This observation aligns with broader trends identified by the ACMA, which found conspiracy theory and health misinformation declined overall in Australia but with notable pockets of persistence. Despite this, the reach of misinformation on Redbubble's platform dropped sharply, with total impressions falling 81% year-on-year. Over 495,000 accounts were disabled globally in 2025 representing a 65% increase on 2024 (including networks of connected accounts responsible for coordinated uploads).

The reports of Apple, Twitch and Redbubble reinforce an important principle for the Code: proportionality. The most appropriate and effective measures for combating misinformation are not uniform across all platforms, but should be calibrated to the platform's specific risk profile, content architecture, and user base. At the same time, all three demonstrate that meaningful transparency reporting is achievable regardless of platform size, with each providing outcome-oriented data such as impressions reduced, enforcements resolved, or user concerns substantiated that allows for public accountability.

Independent Assessment of the 2025 ACPDM Transparency Reports

By: Shaun Davies, Independent Reviewer

My thanks to every signatory for the considerable work behind these reports. Seven were submitted this year. The reports were broadly comparable to last year, with welcome progress on Australian-specific data, generative AI and election integrity. But areas for improvement remain.

A strong focus on election integrity was welcome, though expected in an election year. I also saw a greater focus on generative AI across the board - again, this is no surprise, but it is good to see the reports evolving to encompass this monumental technological shift. It also illustrates a point from my last report: misinformation and disinformation are a constantly moving target, which is why some flexibility in how metrics are reported will always be necessary.

That said, the guidelines ask signatories to provide three years of trended Australian data where feasible. It is a modest request and most signatories provided at least some multi-year or year-on-year Australian data this year. The depth was uneven – a couple gave a full three years, while others compared only the two most recent – but the direction is encouraging.

Meta is the material exception. Its report has some excellent detail and Meta's team was responsive to feedback. But for the second year running it has declined to provide trended Australian tables, even for metrics that appear stable year on year. Meta's position, which it has confirmed to me, is that these figures need additional context and might be misinterpreted without it.

I take that concern seriously and, to be fair, the underlying figures remain in Meta's earlier reports for anyone who wants them. But in my view the best remedy for misinterpretation is a clear explanation for unexpected swings in the data. Readers should not need to self-assemble a comparison from past reports and data points in isolation tell an incomplete story. I do not regard this as a breach of the Code in itself, which does not mandate trended data. But it is a clear departure from the expectations set out in the Transparency Reporting Guidelines.

A related issue across several reports is heavy reliance on repeated language. That is fine where a policy genuinely has not changed, but not where recycled text carries forward stale dates or presents last year's activity as current. The same goes for reused case studies: repeating past successes is not useful to transparency efforts. To their credit, signatories corrected these points readily once I raised them. My concern is that they reached me at all. The review should test a report's substance, not pick up issues that a careful read before submission would have caught.

The current review of the Code is a chance to set practical minimum standards. My view is that future reports should carry a structured Australian data appendix – three years of figures for a defined set of recurring metrics, with plain explanations wherever a metric is considerably altered or dropped entirely. Larger platforms should also include at least two current-year case studies relevant to the reporting period; smaller signatories should be encouraged to do the same where proportionate. I would also welcome more on appeals, reinstatements, corrections, false positives and how platforms weigh enforcement against freedom of expression. This is the harder, more honest side of moderation that these reports still rarely address. My thanks again to all signatories; the direction of travel is generally a good one.

Shaun Davies is the independent reviewer of the 2025 transparency reports

Part 2 | **Code Administration**

This section contains an overview of the key activities of DIGI in its role as administrator of the ACPDM.

Complaints

The Code complaints facility is an important pillar of the Code's governance process which is aimed at ensuring Signatories are accountable for the commitments under the Code including the accuracy of the information in their transparency reports. Eligible complaints can be made by the public, via the complaints portal that DIGI administers on its website, and are escalated to an independent Complaints Sub-committee.

Since the complaints facility opened in December 2021, DIGI has received a total of 91 complaints through the facility (excluding the test responses used during system setup). These complaints have been lodged by members of the public across Australia against signatories to the ACPDM, most commonly directed at Meta (Facebook/Instagram), Google (including YouTube), TikTok, and Twitter/X (a former signatory of the code).

In the reporting period from 1 May 2025 to 30 May 2026, DIGI received 14 complaints. These were spread across a range of signatories, with Meta (Facebook/Instagram) attracting the largest number, followed by Google.

Of all complaints received since the facility's inception, only two were assessed as eligible under the Code's complaints process. Both of these eligible complaints were lodged by Reset Australia, a civil society organisation. The first, submitted in 2024, concerned X (formerly Twitter) and the removal of civic integrity reporting tools and was upheld. The second concerned alleged misrepresentation in Meta's transparency reporting regarding fact-checking labels applied to misinformation content. This complaint was dismissed by the independent committee.

Many complaints received, while reflecting genuine concerns by users, do not fall within the scope of the Code. A significant proportion relate to matters outside DIGI's remit, including:

- **Account suspensions and bans** — several complainants sought DIGI's intervention where their accounts had been suspended or restricted by platforms, often without what they regarded as adequate explanation or appeal pathway. Account-level enforcement decisions are generally not covered by the Code's obligations.
- **Age verification issues** — at least one complaint in the recent period related to an account being suspended pending age verification, with the complainant unable to access the appeals mechanism. Age verification processes fall outside the Code's specific obligations around disinformation and misinformation.
- **Defamation and reputational harm** — a number of complainants raised concerns about content they considered defamatory or damaging to their personal or professional reputation appearing in search results or on social platforms. Defamation is a legal matter rather than a matter of misinformation policy under the Code.
- **Copyright disputes** — several complaints related to copyright strikes applied to content creators' accounts and the failure of platforms to process retractions from rights holders. These are matters of intellectual property enforcement rather than disinformation.
- **Content moderation disagreements** — many complainants objected to platforms removing or restricting their own content, framing this as censorship or a breach of free expression.

Many Australians who interact with the facility are seeking individual redress against a range of platform decisions, rather than raising concerns about how platforms are meeting their obligations to reduce the spread of disinformation and misinformation in accordance with the requirements of the Code. DIGI has made new materials for complainants on its websites that explain how the facility operates in plain language and continues to direct those with out-of-scope concerns to more appropriate channels where possible.

Annual Event

Information Integrity in the Age of AI

In September 2025, DIGI and Meta co-hosted the annual event for the ACPDM focused on information integrity challenges in the generative AI era. This collaborative forum brought together leading voices from academia, industry, government and civil society to foster evidence-based dialogue on Australia's evolving digital literacy needs.

DIGI's Tahlia Davies moderates a discussion on media literacy with Holly Nott, AAP and Professor Tanya Notley, Western Sydney University (WSU)

***The event featured a keynote presentation from** Prof. Nicolas Suzor, Professor at QUT Law School and Oversight Board Member*

Fireside Chat: Driving Media and AI Literacy
Holly Nott (Australian Associated Press) and Professor Tanya Notley (Western Sydney University), moderated by DIGI

Panel Discussion: The Australian Code of Practice on Disinformation and Misinformation
Dr. Michael Davis (University of Technology Sydney), Dr. Anne Kruger (University of Queensland) and industry expert Shaun Davies

Promotion of the ACPDM

DIGI continues to promote key milestones in its governance of the code through media releases and other communications materials. In 2025, DIGI issued two proactive media releases regarding the code, including to support public transparency of the transparency reporting round and the public consultation of the 2025 planned code review. The ACPDM was featured in 10 media stories. DIGI also regularly engages with inbound media requests related to the code to support public transparency and promote deeper understanding of the code's principles, goals, and activities. In 2025, DIGI also undertook a redesign of its website, including transparency, governance and complaints pages, in order to make information about the ACPDM more accessible and clearer for consumers.

DIGI remains active in promoting the code's governance milestones through diverse communication channels and media engagement. To enhance public transparency during 2025, DIGI published two

proactive media releases concerning the annual transparency reporting cycle and the public consultation phase of the 2025 code review, with the result that the ACPDM was highlighted in 10 separate media stories. To improve user experience, DIGI also redesigned its website in 2026, streamlining the presentation of governance, transparency, and complaints information.

The release of this year's transparency reports is timed to align with the Science Misinformation Symposium, an event hosted by the Science Media Centre. This symposium was the result of a strategic collaboration with the *UQ WhatIF Lab*. As a featured speaker at the event, DIGI will share critical findings from the 2025 reports with members of the academic and scientific communities.

Governance Committees

The structural governance of the ACPDM is detailed in Appendix B. Specifically, the Administration Committee is tasked with the following responsibilities:

- Overseeing signatory actions to ensure compliance with Code obligations, including tracking any significant changes made since previous transparency reports.
- Evaluating the performance and utility of the Code Complaints Facility, which involves monitoring the volume and nature of both eligible and ineligible complaints.

Governance committees

The governance arrangements for the ACPDM are set out in Appendix B of this report. The specific functions of the Administration Committee include:

- Monitoring actions taken by Signatories to meet their obligations under the Code, including material changes since their most recent transparency reports.
- Reviewing the operation and effectiveness of the Code Complaints Facility, including the number of ineligible and eligible complaints.
- Reviewing and reporting on Signatories' responses to systemic issues brought to its attention by the Complaints Sub-committee.
- Reviewing and reporting on the effectiveness of the independent review of transparency reports.
- Reporting on progress of relevant research initiatives on misinformation and disinformation.
- Reviewing the annual report produced by DIGI on Code administration.
- Making recommendations to DIGI and Signatories related to issues raised and discussed at meetings.

The primary role of the Signatory Steering Group is to steer the implementation of the Governance arrangements under the ACPDM. Its role includes tasks such as finalising governance arrangements, agreeing on appointments of independent members, determining arrangements for the annual event required under the Code, agreeing on Best Practice Reporting Guidelines, considering changes in government policy, evaluating the need for amendments to the Code, approving DIGI’s Annual Report on Code Administration, and agreeing on the scope of annual reviews of the Code.

During 2025, the Administrative Committee and independent reviewer were consulted on the outcomes of the 2025 Code review, and provided the draft discussion paper and project plan for comment. In November 2025, DIGI held a further update meeting with the independent members of

the Administration Committee to brief them on progress and the analysis of public submissions and emerging recommendations.

Outcome of Code Review

DIGI completed the second review of the ACPDM May 2026. The revised Code came into effect on 30 May 2026. The review was the most comprehensive since the Code's launch in February 2021, examining both the substantive scope of the Code and its governance and oversight arrangements.

Review Process

Background and timing. The second review was originally scheduled to commence in 2024. However, the Signatory Steering Committee agreed to pause the review while the Federal Parliament considered the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024. The Bill would have significantly altered the regulatory landscape for the ACPDM by providing the Australian Communications and Media Authority (ACMA) with formal oversight powers over digital platforms' efforts to address disinformation and misinformation. Following the withdrawal of the Bill, DIGI commenced the review of the Code in September 2025.

Discussion Paper. To guide stakeholder engagement, DIGI published a Discussion Paper on 30 September 2025. The paper set out the background to the review, identified key issues for consideration, and posed specific questions to assist stakeholders in making submissions. A threshold question raised in the Discussion Paper was whether the Code should continue to regulate both disinformation and misinformation, or whether its scope should be narrowed to disinformation only (which is a more objectively verifiable concept). The Discussion Paper also asked for submissions to focus on improvements to transparency reporting, the Code's role in supporting an ecosystem approach to combatting disinformation and misinformation, and the governance arrangements for the Code's administration committee.

Public consultation. DIGI accepted public submissions between 30 September and 3 November 2025. The review drew on 11 submissions from members of the public, academia, civil society, and government, as required under the Code's review provisions. Consistent with section 7.9 of the Code, the review also took into account the ACMA's fourth report to government on the adequacy of signatories' efforts under the Code, as well as the recommendations of the Report by the Senate Select Committee on Information Integrity on Climate Change and Energy which examined the role of social media, bots, and artificial intelligence in spreading climate misinformation.

Outcome. The review concluded that misinformation should remain within the scope of the Code. Signatories agreed that the Code's flexible, opt-in model provided sufficient safeguards to ensure that freedom of speech is protected. Signatories also agreed it was important to keep misinformation in scope of the Code given that disinformation campaigns are often aimed at encouraging and amplifying misinformation.

Key Updates to the Code

The 2026 update introduced several new opt-in outcomes, expanded certain existing provisions, and refined the Code's governance arrangements.

New and updated outcomes. Three new outcomes (1e, 1g and 4a) address the evolving digital environment, particularly the growing role of AI and an updated outcome 1f expands signatories commitment on recommender systems:

- *Outcome 1e (Account Enforcement Transparency)* requires relevant signatories to provide users with information about enforcement action taken against their accounts for violations of misinformation policies. This is designed to improve individual user awareness of platform decisions affecting them and is aligned with the EU Code on Disinformation.
- *Outcome 1f (Recommender Systems)* commits relevant signatories to making information available to users describing how their recommender systems are designed, how those systems influence the visibility and spread of digital content, and what safeguards are in place to limit the propagation of misinformation.
- *Outcome 1g (AI Content Identification)* requires relevant signatories to support users in identifying digital content that has been generated or manipulated by AI systems on their services, for example through labelling or marking of AI-generated content.
- *Outcome 4a (Electoral Integrity)* commits relevant signatories to formal cooperation with federal electoral bodies, including the Australian Electoral Commission (AEC), to promote the integrity of federal elections and increase access to authoritative electoral information. This outcome was directly informed by the Senate Select Committee's concerns about algorithmic amplification and electoral integrity.

Expanded scope on inauthentic behaviour. The Code now explicitly acknowledges that inauthentic behaviour including bot-driven amplification can be used to artificially increase the reach or perceived support for misinformation. This clarification strengthens the Code's existing provisions on platform integrity.

Governance reforms. The administrative structure of the Code was updated to replace the former Administration Sub-committee with a new Independent Advisory Committee. The key change is that the new Committee is composed exclusively of independent members who are not currently employed by the technology industry or its representative bodies. The Committee's role is to provide independent advice to signatories via the Signatory Steering Committee on the operation of the Code, including complaints handling, transparency reporting, and proposed amendments. This change more accurately reflects the Committee's advisory function and is designed to enhance the impartiality of the Code's governance. The updated Code also recognises the ACMA's role in providing informal oversight of the Code and reporting to the government on signatories' efforts (section 7.8).

Enhanced transparency reporting. The Best Practice Transparency Reporting Guidelines have been updated (Version 4.0, May 2026). The updated guidelines introduce a recommended table of more granular data points for relevant signatories with 32 common quantitative and qualitative data points. Signatories will also be required to report against three new outcomes 1e, 1g, and 4a in their transparency reports for the 2026 calendar year.

Complaints facility. The DIGI website has been updated with new materials and clearer guidance on the complaints process, including a plain language explainer and a diagram of the complaints pathway. The Independent Advisory Committee will continue to monitor systemic issues indicated by complaints and may make recommendations to the Signatory Steering Committee for governance improvements.

Appendix A | **Insights from 2026 Transparency Reports on The Australian Code on Disinformation and Misinformation.**

1. Introduction

This report summarises key information from the 2026 transparency reports of the signatories to the Australian Code of Practice on Disinformation and Misinformation:

- Platform Specific Actions and Data Insights (Outcome 1)
- Signatories Commitments and 2025 Policy Updates.

Please note that Signatories will update their commitments for 2026 to include new outcomes introduced in the ACPDM as part of the 2025 review. When available these will be published on the DIGI website.

This report was compiled with the assistance of an AI Tool.

2. Platform Specific Actions and Data Insights (Outcome 1)

Google

Category	Action / Data insights (Jan–Dec 2025)
Content removals (YouTube AU)	136,562 YouTube videos removed for Community Guidelines violations (AU IPs) *889 removed for misinformation or spam, deceptive practices & scams policy violations* *77% removed with 10 or fewer views*
Violative view rate (YouTube global)	Q1: 0.10–0.12% | Q2: 0.14–0.15% | Q3: 0.11–0.13% | Q4: 0.14–0.15%

Category	Action / Data insights (Jan–Dec 2025)
CIB operations	Google Threat Intelligence Group (GTIG) continued to monitor and disable coordinated influence operations globally, publishing quarterly TAG Bulletins. Note *TAG consolidated into a unified GTIG entity during this period.*
AI-related interventions	Launched SynthID Detector portal (to early testers/partners) for third-party verification of AI-generated content across text, audio, and video. Integrated C2PA Content Credentials into Search and YouTube ('Captured with a camera' disclosure). All Imagen 4-generated images are watermarked with SynthID in consumer products. Google–AAP partnership launched to help enhance the usefulness of results displayed in Gemini. 'About This Image' continues across surfaces like Circle to Search and Google Lens; focus shifted to integrating C2PA metadata and SynthID signals in real time.
Recommendation system controls (YouTube)	Users can view, delete, or turn off watch/search history. Users can disable homepage/Shorts recommendations by turning off and clearing watch history. Users have controls to fine tune recommendations.
Ad takedowns (AU)	9,937,100 actions for Misrepresentation Policy violations 1,597,339 actions for Inappropriate Content Policy violations 71,912,499 actions for Destination Requirements violations 151,272 AdSense pages actioned; 146 domains actioned *Gemini-powered tools caught 99%+ of policy-violating ads before serving; 8.3B ads blocked/removed globally; 24.9M accounts suspended globally.*
Ad appeals (AU)	476,145 total appeals *130,558 successful; 0 partially successful; 345,587 failed*
Election ads (AU)	187 verified advertisers ran election ads 37,217 ads rejected (unverified advertisers) Total spend: A$35,808,900 across 27,694 ads

Category	Action / Data insights (Jan–Dec 2025)
	Top states by spend: Victoria (A$11M), NSW (A$9.93M), Queensland (A$6.48M). Video formats dominated at A$32M.
Search transparency features (AU)	'About this result': 65,120,384 views 'More about this page': 3,427,572 views Crisis Response alerts: over 21,600,000 impressions
Information panels (YouTube, AU)	Over 740 million total impressions on information panels in Australia
Human rating / evaluations	Continued use of human evaluators and machine learning for YouTube content moderation and rankings and recommendations. Search Quality Raters help benchmark the quality of results under publicly available Search Quality Rater Guidelines
Media literacy programs	Super Searchers pilot run with South Australia Libraries; curriculum refreshed November 2025 to include AI Overviews and AI Mode. Google partnership with Life Ed: Be Internet Awesome delivered to Australian schools and families. Squiz Kids' Newshounds program: 2,000+ Australian classrooms; 86% of pilot students reported changed media consumption habits.

Source: Google Annual Transparency Report, May 2026

Meta

Category	Action / Data insights (Jan–Dec 2025)
Content removals (AU)	Took action on over 270,000 pieces of content across Facebook and Instagram in Australia for violating Misinformation policies. *Note: Figures reflect an expanded policy scope for the 2025 reporting period — Meta expanded the 'Misinformation' category to include coordinating harm, promotion of crime, and inauthentic behaviour policy violations. Not directly comparable to 2024 figures.*
Content warnings (AU)	Displayed warnings on over 4.8 million distinct pieces of content on Facebook in Australia, based on fact-checking. Displayed warnings on over 236,000 distinct pieces of content on Instagram in Australia (including reshares), based on fact-checking.
Ad removals (AU)	Removed over 9,700 ads in Australia for violating Advertising Standards on Misinformation. Removed over 65,000 ads in Australia for not complying with Social Issues, Elections and Politics (SIEP) ads policy.
Fake accounts removed (global)	From January to December 2025, detected and removed 3.4 billion fake accounts on Facebook. On average, proactively detected and removed over 99% of these accounts before they were reported to Meta. Note Meta does not provide Australia-specific statistics for fake account removals.
AI-related interventions	Rolled out 'AI Info' labels for ad creative images and videos created or significantly edited using Meta's in-house generative AI tools. Labels appear in the three-dot menu or next to the 'Sponsored' label. Applied 'AI Info' labels to organic content when industry-standard AI metadata is detected or when users disclose AI-generated uploads. Required disclosure for SIEP ads containing photorealistic AI-generated or altered images, video, or audio depicting real or realistic-looking people or events. Blocked policy-violating sites from AI training data and Retrieval-Augmented Generation (RAG) processes in response to data-poisoning threat vectors.

Category	Action / Data insights (Jan–Dec 2025)
	Meta continues to collaborate with industry partners, policy stakeholders, and experts to evolve its approach to AI labelling.
CIB operations (global)	Removed 16 new covert influence operations globally in 2025. Q1 2025: Three operations disrupted — originating in China (targeting Myanmar, Taiwan, Japan), Iran (targeting Azeri-speaking audiences), and Romania (domestic). Q2–Q3 2025: Seven operations disrupted — originating in Belarus (targeting Poland), India (domestic), Moldova (domestic), and Russia (targeting Moldova and Sub-Saharan Africa). Q4 2025: Six operations disrupted — originating in China (targeting Taiwan), Iran (targeting Azerbaijan and English-speaking audiences), Pakistan (domestic), and Russia (targeting Sub-Saharan Africa and ~20 countries). Key trends: CIB operations are increasingly using generative AI for content creation; outsourcing to local freelancers; engaging in news media impersonation; and coordinating with each other. Detection remained robust despite evolving tactics.
Fact-checking efforts (AU)	Partnered with two third-party fact-checkers for Australia content: Australian Associated Press (AAP) and Agence France-Presse (AFP). Fact-checked content rated as false or altered is demoted in Feed, removed from Recommendations, Explore, and hashtag pages. Community Notes launched in the US only; no plans to launch in Australia at time of publication.
User controls and transparency	Launched 'Your algorithm' (Tune Your Algorithm / TYA) on Instagram in December 2025 in the US, rolled out to English-speaking markets in January 2026. Allows users to view and edit inferred interests shaping Reels recommendations. Political content controls on Facebook and Instagram — giving users more choice over how much political content is recommended — rolled out globally as of May 2025. Existing controls maintained: hide posts, snooze sources, show more/less, unfollow, manage favourites, report content, 'Not interested' signals. Pages, groups and accounts repeatedly sharing false or altered content are removed from recommendations, have distribution reduced, and lose monetisation and advertising ability.

Category	Action / Data insights (Jan–Dec 2025)
Election integrity (AU)	Ahead of the May 2025 Australian Federal Election, activated Voter Information Units on Facebook and Instagram. This initiative reached approximately 11.8 million Facebook users and 9.5 million Instagram users. Deployed Election Day Reminders on election day, reaching approximately 9.9 million Facebook users and 7.5 million Instagram users. Partnered with the Australian Electoral Commission on voter empowerment products and verified election information. Independent fact-checking for election content provided by AFP and AAP, with warning labels and distribution reduction applied to debunked content. Instagram voting stickers available for users to share civic participation on Stories.
Media literacy programs (AU)	Partnered with AAP to deliver a multilingual media literacy campaign (English, Simplified Chinese, Vietnamese, Arabic) reaching 823,638 Australians, with over 5.07 million views and more than 2.45 million completed video plays. Supported research by La Trobe University and University of Waikato analysing over 3 million posts from 25 major Australian news publishers across 15 years (published September 2025). Hosted hybrid roundtable in Sydney (April 2025): 'Supporting Media Literacy and Information Integrity in the AI Age', convening government, regulatory, academic and industry stakeholders. Co-hosted with DIGI the annual Australian Code of Practice on Disinformation and Misinformation event (September 2025): 'Information Integrity in the Gen AI Era', with over 50 participants. Commissioned research from Queensland University of Technology (QUT) on community-based content moderation (to be published 2026). Supported research by La Trobe University and University of Waikato analysing over 3 million posts from 25 major Australian news publishers across 15 years (published September 2025). Hosted hybrid roundtable in Sydney (April 2025): 'Supporting Media Literacy and Information Integrity in the AI Age', convening government, regulatory, academic and industry stakeholders. Continued collaboration with Australian Strategic Policy Institute (ASPI) on influence operations findings.

Category	Action / Data insights (Jan–Dec 2025)
Research access	Content Library and Content Library API made available to qualified third-party fact-checking partners and eligible academic researchers, providing access to publicly accessible data from Facebook and Instagram.

Source: Meta response to the Australian Code of Practice on Disinformation and Misinformation, May 2026

Microsoft

Category	Action / Data Insights (Jan–Dec 2025)
Content Removals *(LinkedIn, AU)*	LinkedIn removed 933 pieces of misinformation reported, posted, or shared by Australian members LinkedIn globally removed 44,564 pieces of misinformation MSN Australia: 161 misinformation-labelled comment takedowns (0.34% of 47,083 total takedowns), down from 0.77% in 2024 988,892 comments proactively blocked in Australia by MSN automated systems MSN misinformation sub-categories (AU, 2025): COVID-19 – 10 (0.02%); QAnon – 0 (0%); Russia/Ukraine – 4 (0.01%); US Elections – 66 (0.14%)
Fake Accounts Blocked *(LinkedIn, AU)*	LinkedIn blocked more than 1,400,000 fake accounts attributed to Australia in 2025 Globally, LinkedIn blocked more than 197 million fake accounts (majority stopped at registration) AU breakdown (2025): Stopped at registration – 360,721 (Jan–Jun) + 783,615 (Jul–Dec); Restricted proactively – 155,224 + 165,905; Restricted after report – 1,919 + 4,404
Defensive Search Interventions *(Bing, AU)*	Apr–Dec 2025: 324,965 queries and 6,358,885 impressions subject to defensive search interventions in Australia Ukraine-related: 18,726 queries and 248,549 impressions (decrease in impressions reflects reduced user engagement with this topic) Increased queries and impressions correlate to increased use of Bing in Australia

Category	Action / Data Insights (Jan–Dec 2025)
	Interventions include algorithmic ranking adjustments, autosuggest restrictions, and manual interventions for high-risk areas (elections, pharmaceuticals, COVID-19)
AI-Related Interventions *(Bing, LinkedIn, Microsoft-wide)*	Bing partnered closely with Microsoft's Responsible AI team to proactively address AI-related risks in generative AI experiences (Copilot Search, Bing Image Creator, Bing Video Creator) LinkedIn adopted C2PA 'Content Credentials' technology – content containing the "Content Credentials" technology will be automatically labelled on LinkedIn, with users seeing the "Cr" label Microsoft added Content Credentials to all images created via Bing Image Creator, Microsoft Designer, Copilot, and Azure OpenAI DALL-E Microsoft's Reed Smart / Minecraft Education game reaches millions of learners globally (including Australia) on AI misuse detection Search Progress information literacy tool adopted in 10+ countries; updated Q1 2026 with enhanced source-credibility features Microsoft piloted a training for Australian content creators and influencers on AI ethics and deceptive AI (May 2025) Microsoft launched AI & Elections Accelerator reaching 1,000+ election officials from 80 countries
Ad Takedowns *(Microsoft Advertising, Global & AU)*	Globally: ~8.8 billion ads and product offers taken down for policy violations in 2025 (up from 7.9 billion in 2024) Suspended nearly 183,546 customers and blocked 512,442 ads that spread disinformation Australia: 1.7 billion ad rejections; 727,137 total entity takedowns (722,299 in AU) AI-powered advancements used to detect new patterns; coverage extended to text, image, and video content Enhanced integration with Microsoft Threat Analysis Center (MTAC) for Foreign Information Manipulation and Interference (FIMI) domain signals
Ad Appeals & Complaints *(Microsoft Advertising, AU)*	Australia 2025: 77,307 total appeals; 61,416 appeals overturned Total complaints: 1,277 | Policy violations: 823 | Trademark infringement: 434 | User safety: 20 | Other: 0 Average processing time: ~36 hours (consistent with prior years)

Category	Action / Data Insights (Jan–Dec 2025)
	Advertiser Identity Verification (AIV): 7,913 accounts opted in (AU); 6,451 successfully verified
Fact-Checking & Media Literacy *(Bing, MSN, LinkedIn)*	Bing ingests ClaimReview open schema tags to surface fact-check articles in search results, flagging false or unfounded claims with red 'flags' Ongoing partnerships with The Trust Project to fund media literacy campaigns; pro-bono advertising space provided across Microsoft surfaces generating millions of impressions per month MSN expanded dynamic misinformation detection using GPT-based classification, ingesting updated fact-check claims from Reuters and AFP on an ongoing basis MSN migrating comment moderation from GPT-4o to GPT-5 for improved automated detection Microsoft offers free Search Coach app (Microsoft Teams) and AI Foundations learning suite for educators and students
Election Integrity *(Microsoft-wide, AU)*	Microsoft's Elections & Societal Resilience experts met with 150+ people from Australian political parties, newsrooms, academia and government (January 2025) Heightened technical monitoring and threat intelligence deployed for the 2025 Australian federal election via Microsoft Threat Analysis Center (MTAC), Microsoft Security Threat Intelligence Center (MSTIC), Digital Crimes Unit (DCU), and GHOST teams Bing Search directed users asking election-related questions to authoritative sources including the Australian Electoral Commission Microsoft AccountGuard cybersecurity service available to eligible candidates and election authorities in Australia (available in 40 countries)
Political Advertising *(Microsoft Advertising, LinkedIn)*	Microsoft Advertising prohibits all political advertising globally, including ads for candidates, parties, ballot measures, and political fundraising LinkedIn similarly prohibits political advertising, including issue-based ads exploiting sensitive political topics LinkedIn Feed Brand Safety Score: 99%+ safe (Jul–Dec 2025) – measures ad impressions adjacent to content removed for policy violations

Source: Microsoft and LinkedIn Annual Transparency Report, May 2026

Redbubble

Category	Action / Data Insights (Jan–Dec 2025)
Impressions & Clicks on Misinformation Listings *(Redbubble, AU — 2025 vs 2024)*	2025: 9,558 impressions and 195 clicks by Australian users on listings subsequently removed for violating misinformation/disinformation policy 2024 (prior year): 49,496 impressions and 361 clicks — 2025 represents an 81% decrease in impressions and a 46% decrease in clicks year-on-year Works moderated dropped from 832 (2024) to 201 (2025) — a 76% reduction Average time on site before removal fell from 776 days (2024) to 139 days (2025), indicating faster detection and removal of violating content Decline attributed to: continued improvements to proactive detection tooling; routine policy/process updates removing older listings; and reduction in volume of certain high-impression topics
Content Removals – Misinformation Designs *(Redbubble, AU — 2025 by quarter)*	Approximately 100 harmful misinformation moderation actions in 2025 — the highest level since 2022, and up from a low point in 2023 Moderation volumes by quarter: Q1 highest (~35 actions), Q2 (~31), Q3 (~19), Q4 (~17) — declining trend across the year reflecting continued effectiveness of proactive detection and routine process updates Conspiracy theories (QAnon and 9/11-related content) dominated moderation activity throughout all four quarters Medical misinformation moderation remained at consistent levels throughout 2025 Election misinformation remained low throughout 2025 despite the 2025 Australian Federal Election Historical trend: ~325 (2021) → ~110 (2022) → ~25 (2023) → ~65 (2024) → ~100 (2025)

Category	Action / Data Insights (Jan–Dec 2025)
Notable Trends – 2025 *(Redbubble, AU & Global)*	QAnon content experienced a resurgence in 2025, driven in part by Donald Trump's return to the US presidency and ongoing public interest in the Epstein files 9/11-related conspiracy content persisted across all four quarters, consistent with the enduring nature of that topic online Despite the 2025 Australian Federal Election, election misinformation moderation on Redbubble remained low throughout the year
Fake/Inauthentic Account Removals *(Redbubble, Platform-wide — 2025)*	In 2025, over 495,000 accounts were disabled by account abuse detection measures — a 65% increase compared to approximately 300,000 accounts disabled in 2024 Account-level actions include moderation of listings; temporary suspension; account-level warnings; restriction for self-review; account deletion; termination of networks of connected accounts
User Reporting *(Redbubble, AU — 2025)*	In 2025, Australian users submitted 140 reports using the platform's reporting functionality (covering all Community Guidelines concerns, not limited to misinformation/disinformation) Reporting trend on policy violating content (AU): ~200 (2022) → ~260 (2023) → ~90 (2024) → 140 (2025) Reporting mechanism: prominent 'Report Content' link on every product listing page; directs users to a web form where they can describe the concern; reports are submitted anonymously and reviewed by the Content Safety team
Scalable Detection Technologies *(Redbubble, Platform-wide)*	The vast majority of user-generated images and text pass through one or more scalable detection technologies before being accessible on the platform Current detection tools include: Duplicate detection – identifies previously moderated images users attempt to re-upload Image matching – detects content visually similar to images known for disinformation/misinformation Keyword detection – screens text-based fields (titles, tags, descriptions) for terms linked to mis/disinformation Text-in-image matching – spots text-based misinformation embedded within images

Category	Action / Data Insights (Jan–Dec 2025)
	Machine learning – identifies user accounts linked to known policy-violating networks Artificial intelligence – recognises keyword tagging patterns associated with misinformation/disinformation Platform proactively screens for over 350 different content safety topics including incitement of violence, racism, and misinformation Content Safety Team uses unbiased news sources, global health authority reports, and independent fact-checking organisations to inform policy decisions and training content.
Advertising, Monetisation & Recommender Controls *(Redbubble, Platform-wide)*	Violating content is removed and violating accounts are penalised (up to termination), directly disrupting monetisation incentives from artist sales Keyword blocking tools prevent content tagged with misinformation/disinformation-related terms from appearing on offsite marketing platforms where artists generate sales Artists may opt in or out of advertising their products offsite; keyword blocklists prevent products with mis/disinformation-associated titles or tags from appearing on offsite platforms Product recommendations are based on keyword-matching algorithms connecting artist-generated titles, tags and descriptions with user search and navigation behaviour Redbubble has opted out of Objective 5 (Political Advertising) as it does not sell ad space to parties conducting political advertising

Source: Redbubble ACPDM Annual Transparency Report, May 2026

Twitch

Category	Action / Data Insights (Jan–Dec 2025)
User reporting of content concerns	Twitch provides a dedicated, publicly available guide explaining how to file a report, including step-by-step instructions across web and mobile surfaces and best practices for submitting effective reports.

Category	Action / Data Insights (Jan–Dec 2025)
	Users can report potential harmful misinformation actors from multiple entry points including live streams, clips, past broadcasts, and user profiles. Reporting can be submitted anonymously. Users can also contact the Off-Service Investigation Team (OSIT) directly at OSIT@twitch.tv for allegations of severe off-service offences, handled with confidentiality. In Australia during 2025, Twitch received 31 user reports related to misinformation concerns. All reports were reviewed and resolved within 24 hours. In H2 2025, Twitch responded to 94% of all reports globally in under 1 hour, and 99% within 24 hours.
Recommender systems and AI-related interventions	Twitch provides Content Classification Labels (CCLs) so viewers can assess whether a stream contains discussions or debates about politics or sensitive social issues such as elections, civic integrity, military conflict, and civil rights. Content Display Preferences allow users to filter out streams labelled with tags such as Politics and Sensitive Social Issues. These settings apply across recommendations and search results. Content from followed channels may still appear, ensuring users retain access to content they have explicitly chosen to engage with. Users can indicate they are 'not interested' in specific streamers or content categories to inform recommendations, and may block channels to remove content from recommendation surfaces. These preferences can be reviewed and updated at any time through account settings. Twitch provides publicly available information explaining how its recommendation systems operate, including key signals such as watch history, engagement, language, region, and device.
Reducing advertising / monetisation incentives	Twitch's advertising policies prohibit ads containing deceptive, false, or misleading content, as well as political advertising including campaigns for or against a politician, political party, or related to an election. Advertiser-friendly content guidelines link monetisation eligibility to content and account behaviour. Channels with sensitive content categories may receive limited or no advertising demand. Recent enforcement actions for brand safety violations may result in temporary or extended removal of ad eligibility. The majority of ads on Twitch are placed through a managed process where Twitch personnel work directly with advertisers to ensure compliance. Ads placed programmatically through Amazon's Demand-Side Platform (DSP) also undergo review before appearing.

Category	Action / Data Insights (Jan–Dec 2025)
Enabling informed choices and media literacy	Twitch collaborated with media literacy expert MediaWise to develop educational materials teaching streamers and viewers how to better identify and avoid spreading misinformation and disinformation online. These materials are hosted on the Twitch Safety Center. Between January and December 2025, the media literacy materials on the Twitch Safety Center received 1,886 global user visits. Twitch also promoted media literacy resources through creator education initiatives, including a Creator Camp livestream titled 'Media Literacy with MediaWise', featured on Twitch's front page.
Research access and other transparency initiatives	Twitch publishes a biannual Safety Transparency Report outlining how it enforces its Community Guidelines, including the Harmful Misinformation Actor policy. Twitch also provides a publicly available annual transparency report under the EU Code of Practice on Disinformation. Twitch provides open access to its API, which may be used to retrieve most publicly available channel information, such as content classification labels, stream tags, and moderation settings. The API can be used by third-party researchers studying the service. Twitch participates in a range of industry knowledge-sharing initiatives including the EU Code of Practice on Disinformation, the New Zealand Code of Practice for Online Safety and Harms, the EU Hate Speech Code, the EU Internet Forum, and the Global Internet Forum to Counter Terrorism (GIFCT). Twitch recently took an at-large Operating Board seat for GIFCT.

Source: Twitch Transparency Report – May 2026 (ACPDM)

TikTok

Category	Action / Data Insights (Jan–Dec 2025)
Content Removals – I&A Violations *(TikTok, AU — by quarter 2025)*	Proactive removal rate exceeded 97% in every quarter of 2025 – up from consistently above 95% in 2024 Total I&A violation videos removed by quarter (Australia): Q1 (Jan–Mar): 21,845 removed | 99.5% proactive | 69.4% at 0 views | 71.2% within 24hrs

Category	Action / Data Insights (Jan–Dec 2025)
	Q2 (Apr–Jun): 12,687 removed | 99.3% proactive | 84.6% at 0 views | 90% within 24hrs Q3 (Jul–Sep): 10,104 removed | 98.4% proactive | 78.5% at 0 views | 88.5% within 24hrs Q4 (Oct–Dec): 18,965 removed | 97.8% proactive | 92% at 0 views | 94.4% within 24hrs Q1 dip in 0vv/24hr rates attributed to additional sweeps targeting historical violative content ahead of the May 2025 Federal Election; proactive detection remained strong at 99.5% I&A violations represented 1.8%–4.5% of total Community Guidelines violations per quarter in 2025 (up from 1.5%–1.9% in 2024), reflecting an overall decline in total enforcement volume rather than an increase in I&A content All Community Guidelines violations (AU) – total video removals by quarter: Q1: 926,625 | Q2: 716,701 | Q3: 607,935 | Q4: 486,040
Fake Account Removals & Spam/Fake Engagement *(TikTok, AU — by quarter 2025)*	New algorithm rolled out in late Q3 2025 enabled significantly higher removal of fake likes, fake followers and videos from fake accounts in H2 2025 Q1 (Jan–Mar): Fake likes removed: 11,491,846 | Fake likes prevented: 12,447,209 | Fake followers removed: 663,980 | Fake follower requests prevented: 15,516,907 | Fake accounts prevented: 933,296 Q2 (Apr–Jun): Fake likes removed: 12,253,172 | Fake likes prevented: 10,275,730 | Fake followers removed: 1,659,672 | Fake follower requests prevented: 20,573,772 | Fake accounts prevented: 234,745 Q3 (Jul–Sep): Fake likes removed: 31,429,039 | Fake likes prevented: 7,304,360 | Fake followers removed: 12,934,648 | Fake follower requests prevented: 18,501,362 | Fake accounts prevented: 1,206,954 | Videos removed from fake accounts: 555,555 Q4 (Oct–Dec): Fake likes removed: 79,649,643 | Fake likes prevented: 11,084,510 | Fake followers removed: 49,747,938 | Fake follower requests prevented: 18,136,186 | Fake accounts prevented: 688,195 | Videos removed from fake accounts: 427,584 Note: Videos removed from fake accounts data for Q1 and Q2 2025 not available due to data retention policies
Covert Influence Operations (CIO) *(TikTok, Global & AU)*	In 2025, TikTok continued to publish details of all CIO networks identified and removed in its monthly transparency reports

Category	Action / Data Insights (Jan–Dec 2025)
	No CIO networks were identified as specifically targeting Australia in 2025
	Ahead of the 2025 Australian Federal Election, TikTok conducted four rounds of proactive impersonation sweeps across Government, Politician, and Political Party Accounts (GPPPAs)
	Deployed an account creation prevention strategy targeting 473 name variations of high-profile politicians, political parties, and government agencies – contributing to a measurable reduction in new impersonation account creation in the weeks before election day
	TikTok's CIO detection framework assesses coordinated behaviour, identity deception, and attempts to manipulate public debate, drawing on threat intelligence and data science
AI-Related Interventions & AI-Generated Content (AIGC) Labelling *(TikTok, AU & Global)*	Continued C2PA Content Credentials membership (joined May 2024) and Content Authenticity Initiative (CAI) to drive industry adoption of AIGC labelling standards
	In November 2025, announced new AIGC transparency measures including:
	Testing a new 'Manage topics' control to let users choose how much AIGC appears in their For You feed
	Strengthening labelling via creator tools, detection models, C2PA Content Credentials, and invisible watermarking
	Launching a $2 million AI literacy fund to support expert-led content on AI literacy and safety
	AIGC labelling adoption in Australia showed strong growth throughout 2024–2025 (see chart in report for quarterly data)
	Advertisers are required to use the AIGC label for completely AI-generated or significantly modified advertising content
	Recommender system controls updated in 2025 to include: Manage Topics (10+ topic categories); 'Not interested'; Video keyword filters; For You feed refresh
Ad Takedowns *(TikTok, AU — 2025)*	Total ads removed in Australia in 2025: 184,205 (under broader advertising policies covering harmful misinformation risks)
	Breakdown by category:
	Prohibited and Restricted Industry: 55,832
	Prohibited and Restricted Content: 815
	Misleading and False Content: 127,558

Category	Action / Data Insights (Jan–Dec 2025)
	TikTok is in the process of launching granular misinformation-specific advertising data; current figures reflect broader policy categories that encompass misinformation alongside other violations Political advertising is prohibited across all monetisation features; Government, Politician, and Political Party Accounts (GPPPAs) are banned from placing ads and accessing monetisation features
User Reporting Tools *(TikTok, AU & Global)*	In-app reporting interface includes a dedicated 'Misinformation' category with sub-categories: Election misinformation; Harmful misinformation; Deepfakes, synthetic media, and manipulated media Before submitting a report, users are shown what is prohibited under each sub-category to reduce ambiguity and improve report quality Reporting available across all content types: short-form videos, comments, direct messages, accounts, sounds, hashtags, auto-suggestions, LIVE videos, and livestream comments Non-TikTok users can report potentially harmful material via TikTok's publicly accessible website reporting form Users can track report status and view report history under Settings > Help Centre > Safety Centre; users notified of strikes and given opportunity to appeal via the Account Status page
Fact-Checking & Media Literacy *(TikTok, AU — 2025)*	Continued fact-checking partnership with Australian Associated Press (AAP); AAP published up to 29 debunking articles per month on its independently-run website and proactively submitted leads found on TikTok for review TikTok partners with more than 20 International Fact-Checking Network (IFCN)-accredited fact-checking organisations across 130 markets globally Maintained and updated the Harmful Misinformation Guide (available via online Safety Centre), refreshed in 2025 to reflect current policies, definitions, and fact-checking partnerships, including a visual map of global fact-checking coverage Community Partner Channel: 25 Australian partner organisations introduced to the program (including organisations focused on antisemitism, Islamophobia, hate speech and racism); 400+ organisations globally TikTok maintains a database of previously fact-checked claims to assist human moderators in identifying misinformation; content under fact-check review is temporarily ineligible for the For You feed

Category	Action / Data Insights (Jan–Dec 2025)
Election Integrity & Crisis Response *(TikTok, AU — 2025)*	2025 Australian Federal Election (3 May 2025): Deployed coordinated cross-functional election integrity operation spanning policy, trust & safety, content moderation, law enforcement outreach, and advertising integrity Election Centre H5 page (developed with AEC and AAP): received more than 400,000 video views during its active period Election Search Guide received more than 1.9 million views Western Australian State Election: dedicated Search Guide deployed directing users to authoritative sources Safety Advisory Council convened for a dedicated session on election integrity; two expert roundtables held with academics, subject matter experts, and key opinion leaders ahead of the federal election Published enforcement statistics through the Global Elections Integrity Hub during the campaign, providing real-time public visibility into election-related content actions Zero major content integrity escalations during the federal election period Bondi Beach Terror Incident (14 December 2025): Crisis Search Guide activated within hours, linking to official government sources and national mental health support – viewed more than 30,000 times on 15 December
Political Advertising *(TikTok, AU & Global)*	TikTok prohibits all paid political advertising across all monetisation features, including paid ads, branded political content by creators, and other promotional tools Advertisers identified as politicians or political parties are banned at account level from creating advertising accounts or accessing advertising features Limited exceptions: governments and official electoral entities may advertise for public health, safety initiatives, or non-partisan public service announcements Government, Politician, and Political Party Account (GPPPA) designations are published on TikTok's website; these accounts have restrictions applied including no access to advertising or fundraising features

Source: TikTok Annual Transparency Report (ACPDM), May 2026

Apple

<table>
<tr><th>Category</th><th>Action / Data Insights (Jan–Dec 2025)</th></tr>
<tr><td>Curation and credibility signals</td><td>Apple News supports human curation by trained journalists. An editorial team at Apple News vets publishers before they are onboarded to the platform. Outlets are evaluated to ensure they are credible, standards-based, professional organisations. Details on the guidelines are accessible at https://support.apple.com/guide/news-publisher/publishing-on-apple-news-apde42330c66/icloud.

In Top Stories (the most visible part of Apple News), each article is vetted by Australian-based editors who are all experienced journalists. Editors evaluate multiple sources, seeking out stories that are accurate, factual, editorially fair and compelling. Top Stories are not personalised by design; all readers see the same stories.

Apple News has a strict policy on AI-generated content. Articles generated by or with the assistance of AI must be labelled within the News publishing system, ensuring labels appear in the Apple News feed and on individual articles. Publishers are also strongly encouraged to include an explanation detailing how AI was used. Using AI to mislead readers is not allowed and could lead to a publisher's suspension.</td></tr>
<tr><td>Reports of content concerns</td><td>In 2025, Apple News readers worldwide reported approximately 452,000 concerns on articles, covering both technical and content-related concerns.

The vast majority of those reports (approximately 450,000) were not substantiated.

Approximately 2,200 concerns on 800 individual articles worldwide were deemed valid and warranted action from the moderation team. These concerns cover a range of issues and were not limited to misinformation/disinformation.

The following table shows a comparison with numbers from 2021 to 2025:

<table>
<tr><th></th><th>2021</th><th>2022</th><th>2023</th><th>2024</th><th>2025</th></tr>
<tr><td>Concerns reported globally</td><td>655,000</td><td>370,500</td><td>331,000</td><td>495,000</td><td>452,000</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Category</th><th>Action / Data Insights (Jan–Dec 2025)</th></tr>
<tr><td></td><td>
<table>
<tr><td>Concerns deemed valid</td><td>17,000 (5,600 articles)</td><td>6,500 (2,800 articles)</td><td>4,800 (2,500 articles)</td><td>2,700 (1,300 articles)</td><td>2,200 (800 articles)</td></tr>
</table>
In 2025, Apple was able to ascertain the region of origin for approximately 95% of concern reports. Of those, approximately 3.5% came from readers in Australia.
Approximately 52% of reports originating in Australia were categorised by the reader as misinformation or disinformation.
Zero misinformation concerns originating in Australia were deemed valid for that category.
The following table shows a comparison of 2024–2025 data for Australia:
<table>
<tr><th></th><th>2024</th><th>2025</th></tr>
<tr><td>Origin of concern reports from Australia</td><td>4%</td><td>3.5%</td></tr>
<tr><td>Categorised as mis/disinformation by reporter</td><td>53%</td><td>52%</td></tr>
<tr><td>Mis/disinformation reports deemed valid</td><td><5</td><td>0</td></tr>
</table>
</td></tr>
<tr><td>Reducing advertising/monetisation incentives</td><td>
Apple News's design and structure disrupts advertising opportunities for misinformation and disinformation.
Apple's Advertising on Apple News Content Guidelines and Apple Advertising Services Video, Display & Audio Policies prohibit categories of advertisements including ads that are misleading or deceptive.
Apple provides warnings when advertisements violating content guidelines are discovered, and has the ability to block advertisers that repeatedly violate guidelines. Apple does not sell political advertising either directly or through its resellers.
</td></tr>
</table>

Category	Action / Data Insights (Jan–Dec 2025)
Enabling users to make informed choices about source of content	Publishers are identified on each article, in most places with the publication's own branding, so that readers can easily determine the source of their news. Apple employs editors with newsroom experience in reputable Australian journalistic institutions to curate the platform, helping ensure reputable and trusted brands are surfaced to users. Top Stories (approximately 14% of total article views) features only fact-based journalism and is 100% curated by veteran journalists from the Australian news industry who vet each story for adherence to standard journalistic ethics.AI-assisted articles. Publishers are also strongly encouraged to include explanations detailing how AI was used to produce them.
AI-related interventions	Apple makes information about recommendations in News available to users, together with options and tools associated with those recommendations. See https://www.apple.com/au/legal/privacy/data/en/apple-news/. Articles in Apple News generated by or with the assistance of AI must be labelled and marked within the News publishing system, ensuring labels appear in the Apple News feed and on individual articles. Using AI to mislead readers is not allowed and could lead to a publisher's suspension from the platform.
Major news events and election integrity	Australian Federal Election (May 3, 2025): Apple News covered the full election cycle. Coverage was packaged as "Election 2025" and regularly featured in Top Stories. Apple News delivered a real-time results experience built through a data partnership with The Australian, featuring two-party preferred estimates, an interactive seat-by-seat map, and a live seats-won tracker. Bondi Beach Terror Attack (December 14, 2025): Apple News Editors updated Top Stories following verified reports, closely monitoring sources and prioritising accuracy and quality. Only stories incorporating verified reports from NSW Police, emergency services and credible first-hand witness accounts were featured.
Strengthening public understanding	Apple has taken a global approach to supporting media literacy programs, with continued support for The News Literacy Project (NLP), an organisation that empowers young people with critical thinking skills to seek out accurate and reliable information.

Source: Apple Pty Limited, Apple News 2025 Annual Transparency Report, 15 May 2026

3. Signatories Code Commitments and Policy Updates

Signatory	Services with Code Commitments	Code Commitments	Policy Updates (2025)
Google	Google Search, Google News, Google Advertising (Ads & AdSense), YouTube	**Google Search:**1a, 1b, 1c, 1d, 1e , 3, 4, 6, 7 **Google News:** 1a, 1b, 1c, 1d, 1e, 3,4 **Google Advertising** (Google Ads and Google AdSense) 1a, 1b, 1c, 1d, 1e,2, 3, 4, 5,7 **YouTube:** 1a, 1b, 1c, 1d, 1e,2, 3, 4 , 6 ,7	**Google Search:** clarified the site reputation abuse policy. . **YouTube:** Updated Medical Misinformation Policy (COVID-19 and tobacco/nicotine) & Spam, Misleading & Scams to Spam, Deceptive Practices and & Scams Policies.
Meta	Facebook, Instagram including advertising.	All 7 Objectives and Outcomes	**Facebook/Instagram:** Reports changes to third-party fact-checking programmes in the US; Updated penalty protocols; Launched political content controls; Implemented new transparency tools. **Advertising:** Required advertiser disclosure (generative AI in political/social issue ads).
TikTok	TikTok (Global)	All 7 Objectives and Outcomes	**TikTok:** Election Misinformation Policy: updated global policies to address evolving enforcement challenges.
Microsoft	Microsoft Advertising, Bing Search, MSN (Microsoft Start), LinkedIn	**Bing Search**: 1a, 1c, 1d, 3, 4, 7 **MSN (Microsoft** Start): 1a, 1b, 1c, 1d, 4, 7	**Microsoft Advertising:** Critical Events Policy was applied to prevent serving advertising related to the Israel-Hamas conflic s.In 2025, also enhanced its detection

Signatory	Services with Code Commitments	Code Commitments	Policy Updates (2025)
		Microsoft Advertising: 1a, 1b, 1c, 2, 3, 5, 7 **LinkedIn:** 1a, 1b, 1c, 1d, 1e, 2, 3, 4, 5, 7	capabilities through continuous integration with services provided by the Microsoft Threat Analysis Center (MTAC), including signals related to Foreign Information Manipulation and Interference (FIMI) domain
Apple	Apple News	1a, 1c, 1e, 2, 4, 6, and 7	No major changes
Twitch	Twitch	1,2,3,4,6 and 7	Implemented content display preferences. This works by: ● Using **Content Classification Labels (CCLs)** to identify streams that include sensitive or explicit themes.[1] ● Allowing users to filter out streams labeled with specific tags, such as **Politics and Sensitive Social Issues**, across recommendation and search surfaces
Redbubble	Redbubble	1, 2,3,4,6 and 7	No major changes

Appendix B | **Governance arrangements for The Australian Code on Disinformation and Misinformation**

In October 2021, DIGI announced the governance arrangements for the ACPDM in order to strengthen the code and its effectiveness. These are summarised here. The code is a novel self regulatory mechanism that aims to drive improvements through increased transparency about how platforms tackle mis and disinformation; DIGI's governance arrangements have been tailored with that aim in mind.

Complaints committee

The Complaints Committee is independent and resolves complaints about possible breaches by signatories of their commitments under the code. DIGI acts as secretary on this committee, but has no vote on decisions in order to avoid conflicts of interest. The committee meets to hear complaints of material code breaches that cannot be resolved by signatories and complainants. The Terms of Reference for the Complaints Sub-committee can be found on the DIGI website[1], and the three independent members of the Complaints Sub-committee are detailed below.

Administration committee

The Administration Sub-Committee brings together the three independent representatives from the Complaints Sub-Committee with signatories of the code. This committee monitors the various actions taken by signatories to meet their obligations under the Code, such as the operation of the complaints facility. Following the 2025 review this committee will now transition to being exclusively comprised of independent members with an advisory role, with updates to the Code recognising the role of the ACMA in providing oversight of the code and annual reporting.

Signatory steering group

As any digital company can adopt the code, not just DIGI's members, this group enables companies that are not members of DIGI to have an equal say in decisions that are made about the code, if they choose. This group serves to separate DIGI's advocacy work on behalf of its members from the code governance functions.

---

[1] DIGI, *Terms of reference for Complaints Facility and Complaints Sub-committee | The Australian Code of Practice on Disinformation and Misinformation*, https://digi.org.au/wp-content/uploads/2021/10/DIGI-TOR-for-Complaints-Facility-and-Complaints-Sub-committee-_-ACPDM-_-FINAL-NE-1.pdf

Independent review of transparency reports

An independent expert fact-checks all signatories' transparency reports and provides an attestation of them, in order to incentivise best practice and compliance. The reviewer provides advice to the Administration Sub-committee if it cannot provide an attestation of claims in a transparency report.

The attestation process does not involve an evaluation of the quality of the reports or the compliance with the Code, but provides independent confirmation that certain publicly verifiable information is provided in accordance with agreed reporting guidelines. Signatories may also provide an internal contact with whom the reviewer can confidentially verify any internal policies and processes that are not publicly verifiable. The reviewer's role entails:

1. Verifying if each signatory has published and implemented policies and processes that comply with their obligations in sections 5.8, 5.10, 5.11 and 5.13 that pertain to Objective 1 (Safeguards against Disinformation and Misinformation) and Outcomes 1a, b, c and d of the Code. These sections contain the baseline requirements to implement measures that contribute to reducing the risk of users' exposure to disinformation and misinformation, explain prohibited behaviours, provide mechanisms to report disinformation and misinformation, and provide general information on actions taken in response to reports.
2. Verifying if each signatory has published and implemented policies and processes that comply with their obligations in relation to any optional commitments they have made under the Code.
3. Verifying if the policies and processes mentioned in the transparency report are accessible to Australian users.
4. Verification of 1, 2 and 3 involves checking information provided in the transparency report against public sources.
5. Verifying if each signatory is meeting the ACPDM's commitments regarding the form of the reports including the Best Practice Guidelines.
6. Verification will not involve review of sensitive or proprietary information such as the deployment of technological solutions to detect and remove accounts propagating disinformation.
7. Advising each signatory on a confidential basis if they can attest that the report meets these review requirements, or if there are any gaps.
8. Providing advice to the Administration Sub-committee if they cannot provide an attestation in relation to a signatory's reports, in which case the signatory/ies must either amend and resubmit the reports to the reviewer for further assessment or provide written reasons as to why they dispute the reviewer's assessment.
9. Providing a generalised assessment of the reports, which has been included below in this annual report.

Independent Members of Administration Committee and Complaints Committee

Dr. Anne Kruger

Dr Anne Kruger sits on the Complaints Sub-Committee and the Administration Sub-committee.

Anne leads academic and industry collaborative projects aimed at strengthening information integrity. Anne spent nearly four years with global online verification experts First Draft News.

Anne was co-chief investigator and Interim Director at the University of Technology Sydney's Centre for Media Transition which worked with DIGI on the development of Australia's first disinformation and misinformation regulatory code of practice.

A recipient of the UNESCO International Programme for the Development of Communication (IPDC) grant, in 2022 she co-authored a verification and responsible reporting guidebook for practitioners in Southeast Asia.

Anne was an anchor at CNN Hong Kong during SARS, and later a finance reporter at Bloomberg TV. She established an Open Source Intelligence (OSINT) verification lab at the University of Hong Kong collaborating with technologists Meedan, taught news literacy at HKU and led media literacy projects with United Nations Educational, Scientific and Cultural Organization (UNESCO) throughout APAC. She previously held senior editorial, presenter and online positions with ABC Australia and began her career in regional news with Channel Nine's WIN TV. Anne has a PhD in social media verification education.

Victoria Rubensohn AM

Victoria Rubensohn AM is the Consumer Director of Communications Compliance Ltd and Deputy Chair. She served as a Director and Secretary of the Centre for Inclusive Design until November 2025, and as Deputy Chair and Director of ACCAN until late 2024. She served on the International Institute of Communications Advisory Committee until 2025 and is a former Executive Director. Victoria is a Member of the Australian Institute of Company Directors (MAICD).

Christopher Zinn

Christopher Zinn sits on the Complaints Sub-Committee and the Administration Sub-committee.

Christopher has led various successful and disruptive campaigns to help consumers make better decisions in complex markets such as energy, private health insurance and financial services. Christopher heads the www.determinedconsumer.com initiative, is the CEO of the Private Health Insurance Intermediaries Association, sits on the statutory authority reforming the funeral industry, and is on a self-regulatory code committee for the charitable sector. He was also director of communications and campaigns for consumer group CHOICE and has been a reporter and producer for TV, radio and newspapers both in Australia and overseas including the ABC, the Daily Telegraph, Channel Nine, and the UK Guardian.

Shaun Davies

Shaun Davies is the independent reviewer of the 2025 transparency reports.

Shaun Davies is an accomplished digital leader with 20 years of experience spanning content moderation, artificial intelligence, policy, communications, and newsroom leadership. His expertise includes managing global quality and safety for Microsoft Start's content feed, where he set policy and directed cross-functional teams to implement AI systems for moderation.

Shaun is also concluding a Master of Research at the University of Technology Sydney, focusing on the role of journalists in at-scale content moderation for mis- and disinformation. His career reflects a unique combination of practical experience and academic insight into the challenges of the digital age.

Complaints portal

A key component of the governance arrangements is the public complaints portal, that is available on DIGI's website[2]. The operation of the portal is detailed publicly in the complaints facility terms of reference[3], which explains the processes for how complaints are resolved. The resolution measures

[2] DIGI, *Complaints*, https://digi.org.au/disinformation-code/compaints/

[3] DIGI, *Terms of reference for Complaints Facility and Complaints Sub-committee | The Australian Code of Practice on Disinformation and Misinformation*,

have been designed to provide incentives for signatories to address breaches of the code, which is considered a better outcome than more punitive resolution measures.

When a complaint is made through the portal, DIGI assesses its eligibility and escalates the complaint according to a standardised internal process that is overseen and approved by the complaints sub-committee. The complaints form enables members of the public to make complaints where they believe a signatory has breached a code commitment. This approach is consistent with the recommendations of the final report from the ACCC Digital Platforms Inquiry, which recommended an approach to complaints that centres on code breaches through a focus on 'assessing the response of the digital platforms to complaints against the terms of the code'[4].

Signatories to the ACPDM also commit to providing an avenue for the public to make complaints about instances of mis- and disinformation on their platforms. DIGI does not accept complaints about individual items of content on signatories' products or services, and encourages members of the Australian public to report misinformation or materials that violate specific platform policies directly to the code signatories via their reporting mechanisms.

✅ Example of eligible complaint	❌ Example of ineligible complaint
A failure to implement and publish policies and/or reporting that will enable users to report the types of behaviours and content that violates their policies under section 5.10 of the Code.	A determination by a signatory that specific items of content is or is not disinformation, or a decision to remove an individual's account. These are handled under specific reporting mechanisms.

https://digi.org.au/wp-content/uploads/2021/10/DIGI-TOR-for-Complaints-Facility-and-Complaints-Sub-committee-_-ACPDM-_-FINAL-NE-1.pdf

[4] ACCC (2019), *Digital Platforms Inquiry Final Report*, https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf, p. 371

Appendix C | Best Practice Transparency Reporting Guidelines Version 4.0

Prepared for DIGI by Hal Crawford and updated by DIGI following the 2026 Code review

This document provides guidelines for Signatories preparing the transparency reports required by the Australian Code of Practice on Disinformation and Misinformation ("the Code").

The reports themselves fulfill two functions: to inform the public and to provide a framework for the review of activities under the Code.

These guidelines request Signatories provide:

- *Trended data relevant to the Australian market over extended periods*
- *Clear explanations of major changes in policy*
- *Consistency in reported metrics year-on-year*
- *Audience-friendly documents with a minimum of promotional language*
- *Specific information about efforts to combat AI-generated dis/misinformation*
- *A summary of key data points and qualitative information in the form of a table appended to their report.*
- *A summary of key data points and qualitative information in the form of a table appended to their report.*
- *A summary of key data points and qualitative information in the form of a table appended to their report.*

The template follows the same form as the 2022 and 2023 reports, with some additions to reflect the increased scope of the Code from December 2022 and the revisions to the Code made in May 2026.

Introduction

The purpose of this document is to set out guidelines for the annual reporting required of Signatories under the Australian Code of Practice on Disinformation and Misinformation ("the Code"). The annual reports are required under Objective 7 of the Code: "Signatories publicise the measures they take to combat Disinformation and Misinformation."

Signatories filed initial annual transparency reports in May 2021, followed in successive years by reports which have incrementally improved in terms of specificity and consistency.

The general purpose of the reports can be inferred from Objective 7 and the Code's administrative requirements, and is twofold:

- *To communicate to the general public measures taken by Signatories against dis/misinformation*
- *To provide a framework for the independent reviewer, DIGI and other stakeholders to audit compliance with the Code*

Although these aims are related, they could result in significantly different outputs if one function dominated. A key reminder for Signatories is that the documents must be accessible and comprehensible to the general public. The dual purpose of the reports will influence the reporting template recommended in this document.

This is the fourth iteration of these Guidelines following the introduction of Code. This document responds to the expansion of the Code in December 2022, and the revisions to the Code made in May 2026.

Feedback and changes from past reports

Global regulation of dis/misinformation has developed significantly since work on the Australian Code began, and there have been regulatory developments within Australia that may also affect the operation of the Code in future. We focus here on specific requests made to Signatories to improve the utility of their annual transparency reports.

Previous versions of these guidelines asked that signatories:

- *Reduce emphasis on process and policy*
- *Increase use of trended Australia data*
- *Adopt a common reporting period*
- *Use common definitions*
- *Be explicit with objective/outcome commitments*
- *Explain reporting metrics*
- *Provide multi-year metrics reporting*

- *Increase public accessibility*
    - *Observe a word limit (<10,000 words)*
    - *Use breakout case studies*
    - *Use tables, graphs and other visual elements*
- *Reduce promotional tone*

In the 2023 reports two significant additions to the reporting requirements arose from additions to the Outcomes: the introduction of a requirement to provide information about recommender systems (Outcome 1f, previously 1e) and an addition to Objective 2, strengthening the reduction of monetisation incentives for dis/misinformation.

Outcome 1f (previously 1e) gave rise to a reporting obligation to show how signatories have "made information available to the end-user" regarding the operation of recommender engines. Signatories who have committed to this outcome must also include evidence they have made recommender options available to users.

In terms of Objective 2, signatories should speak to their efforts to deter advertisers "from repeatedly placing digital advertisements that propagate Disinformation or Misinformation".

New Outcomes introduced in the May 2026 update which require reporting from 2027 are:

- ***Outcome 1e: Users will be able to access information about enforcement action taken against their accounts by Relevant Signatories in response to violations of policies under 5.10.*** *This requires signatories to provide users with information on enforcement action taken against their accounts for violating policies under 5.10.*
- ***Outcome 1g: Relevant Signatories will support users to identify digital content that has been generated by the use of their AI systems on their Relevant signatories' service.*** *This requires Signatories to implement systems or processes that help users to identify digital content that has been generated by the use of their AI systems on their Relevant signatories' services.*
- ***Outcome 4a: Signatories cooperate with Federal electoral bodies to support the integrity of federal electoral processes.*** *This requires relevant signatories to work cooperatively with Federal electoral bodies (e.g., AEC) to promote*

electoral integrity and increase the accessibility of authoritative information about federal election processes.

Areas for continued improvement are:

- *In the consistent provision of trended data so that comparisons can be made year-on-year*
- *In the provision of data related to the Australian market; and*
- *In clear explanations of relevant internal policy changes.*

In addition, we would like to see discussion and explanation of any measures explicitly undertaken to combat AI-generated dis/misinformation.

Key guidelines

Calendar year reporting

Reports should refer to data from the previous calendar year. The reports filed in May 2027, for example, should relate to the 12 months from 1 January to 31 December 2027 Given that months will have elapsed from the end of that period to the time of compiling the report, it may be acceptable to refer to developments in dis/misinformation in the first half of the current calendar year in passing commentary.

The use of a common reporting period is essential in terms of making comparisons year-on-year and platform-to-platform, and increases the utility of the reports.

Statement of commitments and relevant services/products/platforms

Signatories must state near the beginning of their reports which Code Objectives/Outcomes they have committed to, and which services and products the commitment applies to. This is particularly important for Signatories with big and differentiated portfolios. The omission of a relevant service/product should be noted.

Changes in policy

Signatories should provide details on any significant policy changes related to dis/misinformation and their activities to combat it. This information can be included in the Summary, and in the relevant Outcome sections of the reports. Signatories should clearly explain the change from the old policy to the new, and what prompted the change.

The impact of generative AI

The ACMA has requested Signatories provide information in their transparency reports on specific measures taken to combat dis/misinformation generated by artificial intelligence (AI), which is reflected in the new Outcome 1g. This requirement has led to a structural change to the report template – although the existing Outcomes are engineered to capture activity against this kind of dis/misinformation – but Signatories are directed to consider AI explicitly and provide relevant information under Outcome 1g where available. This also applies to any policy changes that have been instigated by the integration of generative AI on services.

Trended data

In general, there is a need for more trended numerical data in the transparency reports. A minimum of three years of reporting should be supplied with any data, in order to give context. This is a key requirement for understanding.It may not always be possible to supply three years of data for a given metric. In that case, contextualisation through trended monthly numbers may be appropriate.

Accompanying commentary is vital to explain changes. For example, the incidence of detected dis/misinformation may have increased in a given year because the quantum of dis/misinformation increased, or because a Signatory improved detection. Regardless of the potential misinterpretation of trended data, a transparency reporting regime demands it, and furthermore demands that the same data be reported in subsequent years. Any addition or omission of data should also be the subject of an explanatory note.

The expectation of the report review process is that data used as internal Key Performance Indicators in the area of dis/misinformation be included in the report unless there is a clear commercial imperative to omit (see more on KPIs below).

Australian data

Reporting under the Code should provide data for the Australian market. Global metrics may also be relevant, but given the Code's national nature the primary concern should be Australian numbers, examples, and context. It is recognised that Australian data may not always be available: if this is the case, the Signatory should explicitly note that this is the case.

Public accessibility

There are other aspects of the reports that can be built on to improve communication with a general audience:

- *The emphasis on brevity in the first version of these guidelines was perhaps too severe given the big scope of some Signatories' operations. Limiting to 10,000 words should be possible, however.*
- *The use of graphical elements such as bar and line charts helps in communicating numerical information*
- *Breakout (separated from main text body) case studies are recommended to illustrate key points and developments*

Promotional language

One noticeable aspect of some of the first reports was a "promotional" tone. It is natural that Signatories seek to portray their efforts and accomplishments in the best possible light. Unfortunately, promotional language undermines the informational content of the reporting, and encourages cynicism towards what are in fact major and important efforts to curb mis/disinformation. We encourage Signatories to avoid promotional writing and to maintain a neutral stance, highlighting problems and successes with equanimity, and thereby increasing the credibility of the reporting.

We appreciate this can be difficult with a public document: a good rule of thumb is to avoid statements and words that would not be found in internal company reporting.

Generic information

Generic information relating to dis/misinformation process and policy that is unchanged from past reports should be condensed or moved to appendices where appropriate. This is to place greater emphasis on novel aspects of the fight against mis/disinformation, and to avoid losing the novel information among material that is the same year-to-year. The first iteration of these guidelines found that on average, 84% of the content in the initial 2021 reports was generic information relating to dis/misinformation process and policy.

Bearing in mind the dual purpose of the reports – to communicate to the public and demonstrate compliance with the Code – it is necessary to include some of this information repeatedly. For example, it is a mandatory requirement of reporting that Signatories provide links to reporting mechanisms for dis/misinformation (see below for mandatory links). Signatories may also want to include important information about their approach to tackling dis/misinformation in every report, and there should be an opportunity to do this.

Mandatory links reporting

Signatories' commitments under the Code include simple links to information in order that the independent reviewer may assess compliance. To be explicit regarding these mandatory requirements, they are:

- *1b: Links to user guidelines, policies and procedures relating to mis/disinformation*
- *1c: Links to publicly available tools for reporting mis/disinformation*
- *5: Links to/evidence of published information that allows users to better distinguish factual information from mis/disinformation.*

The link requirements are provided as a checklist to ensure simple elements are not omitted. As indicated in the rest of these guidelines, Signatories are expected to elaborate significantly through the identification and provision of relevant data and commentary. Signatories who have not committed to an Outcome/Objective are exempted from the relevant mandatory elements. Note that in addition to providing these links to assist the independent reviewer, it may be helpful for the reviewer to directly query the Signatory on elements of a submitted report.

The issue of KPIs

In the European Union's 2022 Strengthened Code of Practice on Disinformation, great emphasis is put on quantifying the effectiveness of dis/misinformation countermeasures through Service Level Indicators (SLIs) and Qualitative Reporting Elements (QREs) associated with commitments. The idea is that these data may provide a measure of cross-platform comparison. The EU requires that Signatories provide data for the 6-month reporting period on a country-by-country basis.

In practice, the reports filed under the EU Code demonstrate the difficulty in attempting to mandate meaningful shared metrics between platforms that have very different business models, audiences, interfaces and functions. In the three waves of reports filed to March 2024 , the big platforms have created half-yearly reports over 200 pages long with many incomplete tables, often featuring imprecise or missing data. While it appears that great efforts have been made to satisfy requirements, the documents are of questionable use to the public. They do not include trended data – which will accumulate over time as the corpus grows - and need further aggregation and interpretation before they become useful to anyone.We urge the Australian Code's Signatories to identify and commit to appropriate internal KPIs that are consistently reported on from one year to the next.

Challenges in reporting

The Signatories are diverse businesses and there are big variations in the application of the Objectives/Outcomes. It is not possible to be prescriptive in dictating the data supplied in the transparency reports, although it is expected that Signatories themselves identify relevant data and supply it in line with the suggestions of this document (i.e., within the Australian market, for the reporting period, and for a minimum of two years prior to that). A particular challenge may arise for the Signatories whose dis/misinformation operations are extensive. We encourage them to focus on changes within the reporting period and their interpretations, responses and initiatives. There are also some Signatories whose activities relevant to the Code cannot be quantified. In this case, Signatories are encouraged to report case studies and such qualitative information as will increase the general understanding of their efforts.

Note on formatting

We recommend Signatories use their own formatting conventions in terms of font, layout and colour in the final PDF document. This will not hinder independent review and may enhance messaging to the general public. As implied in the template below, the reports should follow a common structure, with considerable leeway for different elements like graphs, tables and breakout case studies. This will ensure a degree of uniformity across Signatories and better enable comparison between reporting periods. Where numerical data can be supplied, the preference is to present this at the beginning of sections and to contextualise with commentary. It is important to clearly explain metrics and the rationale behind the

Annual Transparency Report Structure

We recommended following the framework below in preparing reports. Content suggestions and constraints are given in brackets. Note the positions of graphs and breakout case studies are given as examples only

Summary

[Discuss in brief the overall features of the reporting period]

[Include analysis of the general environment relevant to dis/misinformation]

[Reiterate the primary elements of your work against dis/misinformation]

[Include information here about significant policy changes related to dis/misinformation]

Commitments under the Code

[Use a table to summarise commitments and the platforms they apply to, as below]

1a [paraphrase Outcome 1a]	*[platform] [service] [product]*
1b [paraphrase Outcome 1b]	*[platform] [service] [product]*
1c [paraphrase Objective 2]	*[platform] [service] [product]*
Etc. …	*Etc. …*

[Include short commentary on omitted objectives/outcomes/platforms/services/products]

Reporting against commitments

Outcome 1a: Reducing harm by adopting scalable measures

[Datpoints/Metrics reported and for what reason]

[Comments on trends observed]

[Any changes in type of content/behaviour targeted]

[Changes to acceptable use policy etc.]

[What measures were successful and how is that reflected in the data?]

[Tables and graphics as appropriate]

[Case studies as appropriate]

CASE STUDY 1 *[Illustrates a particular aspect of data trend or impact of changes made] [Note this is an example location for a case study. If appropriate and available, Signatories should provide several case studies. Such qualitative content is valuable in bringing policy to life.]*

Outcome 1b: Inform users about what content is targeted

[What new initiatives in communicating to users what constitutes mis/disinformation?]

[Evidence of user engagement with this content]

[Links to user guidelines, policies and procedures relating to mis/disinformation]

[Note to include information about work against AI generated mis/disinformation]

[Include any policy changes from the last reporting period]

Outcome 1c: Users can easily report offending content

[Any changes in the way users report content for the reporting period]

[Links to publicly available tools for reporting mis/disinformation]

Outcome 1d: Information about reported content available

[What data have you published to users about the amount and quality of dis/misinformation reporting under 1c?]

[Include such data if available]

[Also give links to where the data has been published]

Outcome 1e: information on enforcement action

[What processes are in place to provide users with specific information on why their accounts were subject to enforcement action for violating policies?]

[What information is provided to the user, and how?]

Outcome 1f: Information about recommender engines

[What information have you provided to users about how recommender engines work on your platforms?]

[What options do users have around recommender engines, and how has that been communicated to them?]

[Provide links where possible, or example screenshots if not]

Outcome 1g: support users identify AI

[What systems and processes have been established that enable users to identify if digital content has been generated by the use of AI systems on your service?]

[how are you addressing: labelling/marking of AI-generated content?]

Objective 2: Disrupt advertising and monetisation incentives for disinformation.

[Explain any data points/metrics as above]

[Quantify progress made against the monetisation of disinformation, graphically if possible]

[what measures have been taken against advertisers who repeatedly provide ads containing dis/misinformation?]

[Changes to policies and processes implemented to reduce monetisation for targeted content and behaviour]

[Any relevant changes in market conditions]

Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms.

[Detail of work in the period against inauthentic behaviours that impact product security]

[Note to include information about work against AI generated mis/disinformation]

[As above, detail trends and initiatives, and plans in this area]

[This section may contain reference to 1a, given potential overlap in these Objectives – it is acceptable to simply refer to that section if all actions against inauthentic user behaviour are covered there]

[Include any policy changes from the last reporting period]

Objective 4: Empower consumers to make better informed choices of digital content.

[Detail the ways in which you have helped users distinguish dis/misinformation from quality information]

[What is the uptake or awareness of such "empowerment tools"?]

[In what content categories are they active?]

Outcome 4a: cooperation with federal electoral bodies to support the integrity of federal electoral processes

[Detail cooperation with Federal electoral bodies (e.g., AEC) to promote electoral integrity.]

[What measures have been implemented to increase accessibility of authoritative information about federal election processes?]

Objective 5: Improve public awareness of the source of political advertising carried on digital platforms.

[Detail the ways in which you have flagged political advertising and improved the awareness of political sources of advertising]

[Any challenges on the horizon, e.g. Upcoming elections]

CASE STUDY 2 *[Illustrates a particular aspect of data trend or impact of changes made]*

Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.

[Suggest the use of the table here]

[Name of university/institute/company]	*[Overview of research]*
…	…

...	...

[Notable success/challenges/changes in the above work]

[Include links]

Objective 7: Signatories will publicise the measures they take to combat Disinformation.

[Aside from this report, what other information about your work against dis/misinformation has been communicated to the public?]

[Quantify engagement with this information if possible]

[Overlaps to some extent with 1d, and if there is complete overlap simply refer to that section]

Concluding remarks

[Unanswered questions and challenges]

[Summary of any new initiatives not already mentioned]

[Evolution of your business's understanding of the problem and how to tackle it]

[Observations on the Code and the process of reporting]

[May include developments between the end of the reporting period and now]

Appendix to report

The Appendix to the report should include a summary of quantitative and qualitative data for relevant Signatories At-Risk of Disseminating Misinformation and Disinformation. The following tables present recommended data points for the relevant reporting period.

Part 1: Quantitative Datapoints

This Part 1 data points relating to the Relevant signatories efforts to address misinformation and disinformation. Relevant signatories are asked to provide information that applies to each type of service they offer that is in scope of the Code.

No.	*Information sought for relevant reporting period*
1. Actions on violations	*Does the signatory have data point/s that demonstrate actions taken with respect to content and/or accounts that violate the relevant service/s' policies that prohibit or manage misinformation or disinformation, and is this AU specific or not?*
2. Media literacy	*Does the signatory have data point/s regarding efforts to enable users to critically engage with sources of information on its relevant services e.g labelling, and is this AU specific data or not?*
3. Effectiveness	*Does the signatory have data point/s that demonstrate the effectiveness of relevant service/s' efforts to manage mis and disinformation on its services?*

Part 2: Qualitative data points

Signatories are only expected to include information against the outcomes and measures below that are applicable to their service. This part is intended to supplement qualitative information largely already provided by signatories in their annual transparency reports to provide greater consistency and comparability.

Objective 1: Provide safeguards against harms that may arise from disinformation and misinformation

Code Outcome 1a: Signatories contribute to reducing the risk of harms that may arise from the propagation of disinformation and misinformation on digital platforms by adopting a range of scalable measures .

Policies

No.	*Information sought for relevant reporting period*
4. Disinformation policies	*Does the signatory have policies that prohibit disinformation/inauthentic behaviour on relevant service/s, and if yes, include relevant information including links to relevant documentation, where available?*
5. Misinformation Policies	*Does the signatory have policies that prohibit misinformation, and if yes, include relevant information including links to relevant documentation, where available?*
6. Labelling policies	*Does the signatory have specific policies that restrict (e.g by requiring labelling) artificially produced or manipulated content including AI generated material, and if yes, include relevant information including links to relevant documentation, where available?*
7. Policy changes	*Does the signatory have changes in the service's policies in 1-3, and if so, are the principal changes published?*

Compliance and enforcement

No.	*Information sought for relevant reporting period*
8. Systems and processes for policy compliance	*Does the signatory deploy systems and processes to review user behaviours or user generated content for compliance against the relevant service/s' policies on misinformation or disinformation, and include relevant information including whether human or automated systems and processes or a combination are used with links to relevant documentation, where available?*
9. Human resources	*Does the signatory have dedicated human resources to review UGC content for compliance against the relevant service/s' policies on misinformation or disinformation, and if yes, include relevant information including links to relevant documentation, where available?*
10. Enforcement	*Does the signatory have systems and processes that set out how it enforces compliance by end-users with the relevant service/s' policies on misinformation or disinformation, and if yes, include relevant information including links to relevant documentation, where available?*
11. Promotion of reliable sources	*Does the signatory take steps to actively recommend/promote reliable sources of content to active Australian end-users of the relevant service/s, and if yes, include relevant information including links to relevant documentation, where available?*

No.	*Information sought for relevant reporting period*
12. removing/demoting/downranking of violative content	*Does the signatory take steps to remove/demote or downrank content that violates the service's policies/terms of service concerning disinformation/misinformation, and if yes, include relevant information including links to relevant documentation, where available?*
13. Human rights	*How does the relevant signatory promote human rights in implementing code commitments?*

Code Outcome 1B: Users will be informed about the types of behaviours and types of content that will be prohibited and/or managed by Signatories under this Code.

No.	*Information sought for relevant reporting period*
14. End-user information	*Does the signatory publish information that is accessible to Australian end-users, about the types of behaviours and types of content that will be prohibited and/or managed by the signatory under the Code?*
15. Notification of action against users' accounts	*Does the signatory notify active Australian end-users of the service when action is taken against their account, or content they publish on the relevant service/s, for violating the service's policies that prohibit or manage disinformation and misinformation, and if yes, include relevant information including links to relevant documentation, where available?*

No.	*Information sought for relevant reporting period*
16. Review of account action	*Does the signatory allow active Australian end-users of the service to seek a review of action taken related to the violation of the service's policies that prohibit or manage disinformation and misinformation, and if yes, include relevant information including links to relevant documentation, where available?*
17. Fact checking	*Does the signatory invest in fact-checking e.g (partnerships with fact-checking organisations) or other types of collaborative partnerships that aim to support its efforts under the ACPDM, and if yes, include relevant information including links to relevant documentation, where available?*

Code Outcome 1C: Users can report content or behaviours to Signatories that violate their policies.

No.	*Information sought for relevant reporting period*
18. Reporting of policy violations	*Does the signatory allow all Australian users of the service to report user-generated content/or on-platform activity for violating the relevant service/s' policies on misinformation/disinformation, and if yes, include relevant information including links to relevant documentation, where available?*

No.	*Information sought for relevant reporting period*
19. Scope of reporting options	*Does the signatory provide reporting options available to active-Australian end-users that cover all content available on the relevant service/s (e.g both UGC and advertising)?*

Code Outcome 1D: Users will be able to access general information about Signatories' actions in response to reports.

No.	*Information sought for relevant reporting period*
20. Information about response of Signatory to reports	*Does the signatory give active Australian end-users access to general information about the signatory's actions in response to reports about the relevant service/s, and if yes, include relevant information including links to relevant documentation, where available?*

Code Outcome 1E: Users will be able to information about Signatories' actions in response to violations of their policies.

No.	*Information sought for relevant reporting period*
21. User information about account actions	*Does the signatory provide users with information where their accounts on relevant service/s have been subject to enforcement action for violating mis/disinformation policies, and if yes, include relevant information including links to relevant documentation, where available?*

Code Outcome 1F: Users will be able to access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.

No.	*Information sought for relevant reporting period*
22. Recommender systems design and operation	*Does the signatory give active Australian end-users access to general information about the signatory's design and operation of recommender systems on relevant service/s, and if yes, include relevant information including links to relevant documentation, where available?*
23. Customisation options on recommender systems	*Does the signatory provide active Australian end-users with in-service options to customise the content suggested by recommender systems (e.g to restrict content suggestions), and if yes, include relevant information including links to relevant documentation, where available?*

No.	*Information sought for relevant reporting period*
24. Tools to block/ mute content	*Does the signatory offer active Australian end-users tools to block or mute content posted by other accounts on the service, and if yes, include relevant information including links to relevant documentation, where available?*
25. Options to users to make changes to recommended content	*Does the signatory offer options in relation to content suggested by recommender systems that enable active Australian end-users to make changes based on topics, themes or narratives suggested by recommender systems used by the service, and if yes, include relevant information including links to relevant documentation, where available?*

Code Outcome 1:G Relevant signatories that enable the generation and dissemination of AI generated and manipulated content as part of their service will support users to identify digital content on their service that has been generated by the use of their AI systems

No.	*Information sought for relevant reporting period*
26. AI systems on signatory's services	*Does the signatory develop or operate AI systems that disseminate AI generated and manipulated content through any of its relevant services?*

No.	***Information sought for relevant reporting period***
27. Assistance to users to identify AI manipulated content	*Does the signatory take steps to assist users identify digital content that has been manipulated by AI systems on their relevant service/s e.g requirements for use of AI systems such as labelling or marking, and if yes include relevant information including links to relevant documentation, where available?*

Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms

Code Outcome 3: The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.

No.	***Information sought for relevant reporting period***
28. Integrity and security of services	*Does the signatory have measures in place on the service which prohibit the types of user behaviours that are designed to undermine the integrity and security of the relevant service/s (inauthentic user behaviour), and if yes, include relevant information including links to relevant documentation, where available?*
29. Types of inauthentic behaviours signatory acts against	*Does the signatory list the inauthentic user behaviours acted against in Australia such as fake accounts, bot-driven amplification, or artificial reach for disinformation, and provide relevant information including links to relevant documents, where available?*

No.	***Information sought for relevant reporting period***
30. Actions for inauthentic behaviour policy violations	*Does the signatory take actions against content or end-users/accounts that violate their policies concerning inauthentic behavior, and include relevant information including links to relevant documentation, where available?*

Part 3: Additional Reporting

Objective 7: Signatories will publicise the measures they take to combat Disinformation.

Outcome 7: The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation

No.	***Information sought for relevant reporting period***
31. Additional reporting of inauthentic behaviors	*Does the signatory provide additional reporting steps taken to tackle disinformation e.g reports under the EU code on actions taken against bot-driven amplification, fake accounts, and deep fakes?*
32. Ad hoc reporting	*Does the signatory provide any additional ad hoc reporting in times of crisis e.g during wars such as the war in the Ukraine?*