Digital platforms' efforts under voluntary arrangements to combat misinformation and disinformation
Third report to government

SEPTEMBER 2024

Canberra
Level 3
40 Cameron Avenue
Belconnen ACT

PO Box 78
Belconnen ACT 2616

T +61 2 6219 5555
F +61 2 6219 5353

Melbourne
Level 32
Melbourne Central Tower
360 Elizabeth Street
Melbourne VIC

PO Box 13112
Law Courts
Melbourne VIC 8010

T +61 3 9963 6800
F +61 3 9963 6899

Sydney
Level 5
The Bay Centre
65 Pirrama Road
Pyrmont NSW

PO Box Q500
Queen Victoria Building
NSW 1230

T +61 2 9334 7700
F +61 2 9334 7799

Contents

Executive summary

Australians continue to be concerned about misinformation on digital platforms, with levels of concern among the highest in the world.

Recent data indicates that 75% of Australians are 'concerned' about misinformation and disinformation (referred to collectively as misinformation) – an increase from 69% in 2022.[1]

Most recently in Australia, these concerns have focused on misinformation about the identity of the killer of 6 people in the Bondi Junction Mall in March 2024. In the UK, recent misinformation about the nationality of the killer of 3 young children in Southport directly led to violent protests and attacks on the nation's streets.

The ACMA continues to monitor digital platforms' activities under current self-regulatory arrangements.

The Australian Communications and Media Authority (ACMA) oversees the [Australian Code of Practice on Disinformation and Misinformation](#) (the code), which is administered by the Digital Industry Group Inc. (DIGI).

A key aspect of our oversight role is to promote and support transparency about the actions signatories take to meet their code commitments. One of the ways we achieve this is by reviewing the annual transparency reports produced by code signatories and reporting on their progress towards achieving the outcomes contained in the code, as well as developments over the previous year, to government.

This is our third report to the Australian Government on the effectiveness of the code and signatories' reporting activities. Our previous reports have outlined concerns with the quality of reporting under the code and recommended that the ACMA be provided with regulatory powers to intervene if code administration arrangements prove inadequate, or the code fails.

Transparency reports continue to provide some qualitative insights into signatories' responses to misinformation.

In the period covered by this report, platforms focused efforts on the resilience of their systems and processes, directed users to authoritative sources of information and undertook initiatives to manage misinformation about the Aboriginal and Torres Strait Islander Voice to Parliament Referendum (the Voice Referendum) in October 2023.

It was positive to see signatories use case studies to highlight their interventions for topics of particular social significance. Half of the signatories in this reporting period (Apple, Google, Meta, Microsoft and TikTok) provided data and qualitative commentary on the changes to their systems and processes, or new measures they implemented, ahead of the 2023 Voice Referendum.

Some signatories (Google, TikTok and Meta) also addressed feedback from the ACMA's July 2023 report to take further steps to adapt quickly to technology and service changes by reporting on how they have addressed AI technologies. The code's reporting guidelines were also updated to encourage signatories to provide this information.

---

[1] University of Canberra, Digital News report: Australia 2024, [APO website](#), 2024, accessed 21 August 2024.

However, transparency reports are inconsistent, data is patchy and Australians cannot be confident that platforms are delivering on code commitments.

Since our last report to government in July 2023, minimal if any improvement to qualitative reporting has been identified in the latest round of transparency reports. We consider signatories need to report data against each outcome it has committed to. Without this data, it is difficult to track the progress of measures against the relevant outcome. We consider signatories should be reporting more trended Australian data in their report. With the reports in their fourth year, it is disappointing to see limited progress in this area.

Our detailed data analysis has also identified data integrity issues. While signatories have put in significant effort to collect and report data annually, across the board, signatories do not appear to have a consistent approach to reporting data year-on-year.

There has also been no discernible progress towards identifying key performance indicators by both individual signatories and across the code

While acknowledging the difficult task of developing industry-wide metrics, it is disappointing that signatories have not made progress towards identifying internal key performance indicators (KPIs) for data they already report year-on-year. The lack of progress was also observed by DIGI's independent transparency report reviewer. Our previous reports have consistently emphasised the importance of developing and reporting consistently against KPIs for the success of the code.

The ACMA has reinvigorated work on a code measurement framework and will work with industry to implement the next iteration in the next cycle

With the limited progress in improving the code's reporting framework, in early 2024 we reinvigorated work on a measurement framework to bolster the code's reporting framework and improve transparency about digital platforms' measures to combat misinformation.

Since 2020, we have reinforced the message that to be both effective and enforceable, the code must be supported by robust governance and reporting frameworks. We have stressed the need for signatories to report consistent Australian-specific data points, with associated KPIs to measure the effectiveness of their systems and processes.

These reporting elements are vital to enable the ACMA and the Australian public to assess whether the code is providing adequate community safeguards against misinformation, while also protecting freedom of expression. Our analysis of the latest transparency reports reinforces the need for a robust reporting framework.

Internationally, governments are demanding more transparency from digital platforms and establishing frameworks to hold them accountable.

The implementation of regulatory responses to online harm internationally is starting to impact digital platform behaviour. The European Union and the United Kingdom have developed*,* and are implementing*,* novel regulatory approaches to these issues. Given the global nature and influence of digital platforms, governments are sharing information and experiences about their work to address misinformation and aligning, where possible, their respective approaches.

The ACMA's 2021 recommendations to government on transparency remain relevant and their implementation is more urgent than ever, given community concern.

The Australian Government is also working towards introducing new legislation shortly to grant the ACMA powers and functions in relation to digital platforms. The legislation, if passed by the parliament, would enable the ACMA to collect information to improve the transparency of platforms, including those that are not signatories to the voluntary code.

This would significantly improve current arrangements and provide greater confidence to Australians that misinformation was being actively addressed by digital platforms.

Introduction

In Australia, minimising the risk of harm from online misinformation and disinformation (referred to collectively as misinformation, unless disinformation is specifically stated) on digital platforms has been the subject of self-regulation since 2021. The [Australian Code of Practice on Disinformation and Misinformation](#) (the code) currently has 9 signatories: Adobe, Apple, Google, Meta, Microsoft, Redbubble, TikTok, Twitch and Legitimate.[2]

The code was developed by DIGI at the request of the Australian Government and requires signatories to identify, assess and address misinformation on their services. Signatories must have systems and processes in place and actively use a range of measures to address misinformation and the inauthentic behaviours that may sit behind it. DIGI administers the code and governance arrangements to encourage compliance.

On 31 May 2024, DIGI published transparency reports for the 2023 calendar year on its [website](#) along with its annual report, which includes a statement on the transparency reports by its independent reviewer.[3] These reports are produced annually, setting out signatories' progress towards meeting relevant outcomes contained in the code. Information about the signatories' current commitments under the code is at Appendix A.

Our role to support and promote transparency under voluntary arrangements

In December 2019, the ACMA was tasked by government with overseeing the code's development and, since that time, we have reported to the government on its effectiveness. To date, we have provided 2 reports following assessment of the signatories' transparency reports.

In our first report ([the June 2021 report](#)), we recommended that the ACMA be provided with formal information-gathering powers (including powers to make record-keeping rules) to incentivise greater transparency by digital platforms. We also recommended that we be given reserve powers to register industry codes, enforce industry code compliance and make standards relating to the activities of digital platforms' corporations. These powers would provide a mechanism for further intervention if code administration arrangements proved inadequate, or if the voluntary industry code failed.

In our second report ([the July 2023 report](#)), we provided commentary on revisions to the code following DIGI's 2022 code review. We also made observations on the effectiveness of voluntary arrangements, including that:

- industry needs to take further steps to review the scope of the code and its ability to adapt quickly to technology and service changes
- there remains an urgent need to improve the level of transparency about the measures platforms are taking to address misinformation and their effectiveness
- better reporting by signatories is needed to enable an assessment of progress and impact
- complaints and code administration arrangements still need further development.

---

[2] Twitch became a signatory in February 2024. Legitimate also became a signatory in May 2024, and therefore will provide its first transparency report in May 2025 for the 2024 calendar year.

[3] DIGI, [Australian Code of Practice on Disinformation and Misinformation - Annual Report](#) [PDF], DIGI website, 2024, accessed 19 June 2024, p. 9.

Flowing from these observations, the ACMA reinvigorated work on a measurement framework to bolster the code's reporting framework and improve transparency about platforms' measures to combat misinformation. The measurement framework is discussed in greater detail in this report and at Appendix B.

Australian Government progress to introduce new rules

On 25 June 2023, the Minister for Communications, the Hon Michelle Rowland MP commenced consultation on an exposure draft of the [Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023](#) (the draft Bill) for public comment. If passed by parliament, the draft Bill would introduce a suite of graduated powers for the ACMA to act should voluntary arrangements prove ineffective. The powers and functions proposed in the draft Bill reflect the ACMA's recommendations. The government has consulted extensively on the draft Bill and is expected to introduce new legislation shortly.

Objectives of this report

As part of our ongoing oversight role, it is timely to report to government on:

- developments since our last report to government
- our analysis of the latest transparency reports
- our views on code governance
- next steps for the 2024–25 financial year.

Digital platforms' activities since ACMA's July 2023 report

Since the ACMA's July 2023 report to government, misinformation has continued to impact matters of public interest, democratic processes and conflicts across the world.

Most recently in Australia, these concerns have focused on misinformation about the identity of the killer of 6 people in the Bondi Junction Mall in March 2024. In the UK, recent misinformation about the nationality of the killer of 3 young children in Southport directly led to violent protests and attacks on the nation's streets.

Digital platforms have responded to the changing risks from misinformation on their services

In Australia, platforms focused efforts on the resilience of their systems and processes, directed users to authoritative sources of information and undertook initiatives to manage misinformation about the Aboriginal and Torres Strait Islander Voice to Parliament Referendum (the Voice Referendum) in October 2023.

The increasing use of generative artificial intelligence (Gen AI) has made it easier for bad actors to create and disseminate misinformation at scale, exacerbating the difficulties in determining whether content is artificially generated. Gen AI continues to play an influential role in the propagation of inauthentic behaviours. In 2024, with around half of the world's population voting in elections, tackling this issue remains a priority.

The COVID-19 pandemic highlighted misinformation as an issue capable of posing harm to individuals, businesses and democratic institutions. As the prominence of the pandemic has subsided, many platforms have and continue to reconsider their policies around COVID-19 related misinformation. For example:

- X no longer enforces its COVID-19 misleading information policy.
- YouTube instituted a long-term policy framework to address a range of topics related to health misinformation such as COVID-19, vaccines and reproductive health.
- Meta will continue to remove COVID-19 misinformation that is 'likely to directly contribute to the risk of imminent physical harm', following the requested review of the approach to harmful COVID-19 misinformation by its independent Oversight Board.

Digital platforms are also developing new policies to address the influence of emerging technologies like Gen AI. Many are developing their own Gen AI models to diversify their products and to enhance user experience on their platforms. At the same time, platforms are both developing and revising their systems and processes to address the increasing manipulation of their services by Gen AI as well as to detect and identify content that is generated by AI to increase transparency for users. Part 3 of this report expands on some of the initiatives adopted by digital platforms related to Gen AI.

Governments are demanding more transparency from digital platforms and establishing frameworks to hold them accountable

Digital platforms continue to refine their policies and practices, seeking to protect individuals and society from related harms and to strengthen their systems to combat the evolving tactics of bad actors. However, in many parts of the world, governments are calling on digital platforms to do more to address misinformation on their services, and to be more transparent about the effectiveness of their efforts.

There has been no single global approach to combatting misinformation, and in all cases, this requires a multi-faceted approach. Some jurisdictions enforce legislation aimed at improving transparency and/or to criminalise the spreading of misinformation. Others implement a range of non-legislative means, such as media literacy and public education campaigns, to build resilience of their citizens, while many implement a combination of both, involving a range of stakeholders to establish a 'whole of society' approach to combat misinformation.

While many jurisdictions remain in the developmental phase of online harms regulation, the European Union and United Kingdom have designed novel approaches to mitigating the harms caused by misinformation. Given the global nature and influence of digital platforms, it is important that governments share information and experiences about their work to address misinformation to pursue, where possible, alignment between their respective approaches.

European Union
The European Union's *Digital Services Act* (DSA) places a range of obligations on very large online platforms and search engines (VLOPs/VLOSEs), including transparency reporting obligations and to identify and implement mitigation measures to address significant systemic risks related to, among other matters, public security and electoral processes.

While the DSA is still a relatively new piece of legislation, it appears to have already had a meaningful impact on how platforms address online harms. [A report by the European Commission](#) on the impact of the legislation found, among other matters that:

- X, Apple, Facebook, Instagram and TikTok have all made it easier to report illegal content on their services in response to the DSA requiring platforms to put in place measures to counter the spreading of illegal goods, services or content online.

- The DSA Transparency Database has also provided more visibility about content moderation decisions by providers of online platforms accessible to the public.

- The DSA requires VLOPs and VLOSEs to identify, analyse and mitigate risks related to the electoral processes and civic discourse, while ensuring protection of freedom of expression. This requirement has resulted in a range of improvements, including clearer escalation processes of disinformation and misinformation and an increase in fact-checking capabilities.

The EU's 2022 Strengthened Code of Practice on Disinformation (the strengthened EU code) remains a voluntary code – it plays an important role in addressing misinformation. On 29 July 2024, the European Board for Digital Services, with the EU Commission, [published a post-election report](#) on the EU elections of June 2024, which calls on signatories to the strengthened EU code to request its conversion into a code of conduct under the DSA's co-regulatory framework.

A range of online platforms are signatories to the strengthened EU code, and VLOPs/VLOSEs can sign up to the code as a risk mitigation measure, which includes publishing transparency reports under a comprehensive, world leading reporting framework about how they are meeting their obligations under the code.

The framework establishes a three-tiered approach to measuring the effectiveness of the strengthened EU code: qualitative reporting elements, service-level indicators and structural indicators.

Qualitative reporting elements require signatories to provide qualitative information about the nature and impact of their actions toward fulfilling their commitments under the code. Service level indicators are quantitative data points signatories provide on the impact and effectiveness of specific measures adopted under the code.

Structural indicators go above the service-specific indicators and are intended to represent cross-industry metrics that measure the overall impact of the code in addressing the harms caused by misinformation for each relevant signatory, as well as the entire online ecosystem in the EU, and at the Member State level. Commitment 41 of the strengthened EU code requires signatories to establish a working group and work with a taskforce to develop structural indicators.[4]

The European Commission published the [pilot analysis in September 2023 (developed by TrustLab)](#) comprising of a case study based on 2 structural indicators – prevalence and source of disinformation, across Facebook, Instagram, LinkedIn, TikTok, Twitter (now X) and YouTube in Poland, Slovakia and Spain. In the publication, the working group and the European Digital Media Observatory (EDMO) noted that due to the comprehensiveness of its proposal, the limited time available and data constraints, they concentrated their focus on a pilot analysis.

The EU's experience in developing and implementing a multi-pronged reporting framework to measure the impact of platforms' efforts to address misinformation has informed the ACMA's approach to developing its own measurement framework. Part 2 of this report expands on the development and objectives of the ACMA's draft measurement framework.

United Kingdom
The United Kingdom Government passed the *Online Safety Act 2023* (OSA) on 26 October 2023. The OSA places a new duty of care for online platforms to remove illegal content (including foreign interference) and take down material that breaches their terms of service. Under this approach, platforms will only be obliged to take down misinformation if it violates their terms of service.

Ofcom, the regulator responsible for enforcing the OSA, is undertaking a phased approach to its implementation, with the first phase of duties on illegal content to take effect at the end of 2024. It is currently undertaking public consultation on its draft codes of practice for illegal content, which involves combatting child abuse, fraud and terrorism (including protecting against foreign interference).

The second phase of implementation will focus on child safety and protecting women and girls from harmful online content.

---

[4] The 'taskforce' is composed of representatives of signatories, the European Regulators' Group for Audio-visual Media Services, the European Digital Media Observatory and the European External Action Service and is chaired by the Commission.

As part of phase 3, Ofcom must produce a register of categorised services which will need to produce transparency reports and operate in line with their terms of service. Ofcom has commenced consultation on draft transparency guidance.

Ofcom is in the process of establishing an Advisory Committee on Disinformation and Misinformation that will give advice on how providers of regulated services should counter disinformation and misinformation.

The ACMA's oversight of the code

Since our July 2023 report, we have undertaken initiatives to fulfill our oversight role. We have also conducted ongoing monitoring activities to stay across misinformation-related developments in Australia and globally.

Engaging with industry on their efforts to address misinformation

In late 2023 and early 2024, many digital platforms responded to misinformation about the Voice Referendum and the stabbing events in the Sydney suburbs of Bondi Junction and Wakeley. Online misinformation affected the quality of information reported about and circulating across digital platforms and mainstream media about both events.

In February and May 2023, the ACMA wrote to a range of signatories and non-signatories to the code (Reddit, Spotify, Snap Inc. Yahoo, Discord, Pinterest, Telegram, Tumblr, Twitch and WeChat), seeking information about their efforts to address misinformation during the Voice Referendum and took the opportunity to reiterate previous encouragement that they sign up to the code. We received responses from Reddit, Spotify, Snap Inc, Tumblr, Twitch and Yahoo, which primarily related to how existing policies would address any misinformation related to the Voice Referendum. Some platforms also discussed engagement with the Australian Electoral Commission to support the referendum process.

On 17 April 2024, to understand how digital platforms were addressing misinformation on their services during the Bondi Junction and Wakeley Church attacks in Sydney, the ACMA requested information from code signatories and major non-signatories Snap Inc, X Corp, Yahoo, Reddit, Discord and Telegram, about their responses to misinformation following these events.

We received responses from Adobe, Apple, Discord, Google, Meta, Microsoft, Reddit, Snap Inc, Tiktok, Twitch, and Yahoo with varying degrees of information about their efforts to address misinformation about these events. The 'Next steps' part of this report expands on the nature of these responses.

New signatories, Twitch and Legitimate, sign up to the code

In addition to communicating with platforms about their efforts to address misinformation, the ACMA encourages platforms to participate in the code. While X was removed from the code for non-compliance,[5] it was positive to see Twitch – an interactive live-streaming service – and Legitimate – an AI-tool for journalists – sign up to the code in 2024. Twitch published its first transparency report in the May 2024 reports, outlining its commitments to the code and providing some qualitative and quantitative information about how it is meeting those commitments. The ACMA expects Legitimate will publish its first transparency report in 2025.

Twitch's commitments under the code
All outcomes except Outcome 5 (Twitch does not allow political ads on its platform)

---

[5] Complaint by Reset Australia against X (f.k.a Twitter) upheld by Australian Code of Practice on Disinformation and Misinformation independent Complaints Sub-Committee | DIGI

The ACMA progresses work to develop a measurement framework

The ACMA sees a key aspect of its role in overseeing the code as encouraging DIGI and signatories to transparently report on the measures taken by signatories under the code. This includes highlighting how industry can achieve better performance under the existing code.

There is a lack of observable progress by industry towards an industry-wide measurement framework. This has led the ACMA to start developing a framework to provide a holistic representation of how effectively signatories are fulfilling their obligations through the systems and processes they have put in place.

We have developed a draft measurement framework that:

- provides greater transparency on the systems and processes signatories have put in place to meet code outcomes
- establishes a set of industry-wide metrics to collect quantitative data on the impact of signatories' systems and processes over time
- enables signatories to develop their own internal KPIs against those metrics
- allows for greater 'like-for-like' comparison of signatories' efforts under the code where appropriate.

Our work has been informed by the approach under the strengthened EU code. We recognise that the landscape for digital platforms is complex, as they offer a diverse range of services and operate under disparate business models. As is evidenced by the EU's experience, this creates a challenging environment for developing a measurement framework which is 'industry-wide'. Given the novel nature of the framework and the related challenges, we have engaged in a multi-staged, targeted consultation process across government, with international stakeholders and industry.

Given the technically complex undertaking of developing a measurement framework, the ACMA procured independent expert advice from Moonshot CVE Ltd (Moonshot) to conduct a validation exercise on the framework. Its full report is published alongside this report on the ACMA website.

Moonshot concluded that there are significant challenges to third-party verification of reported data or independent data collection, making it essentially infeasible. Noting these third-party verification challenges, under current voluntary arrangements, the success of the framework to provide necessary transparency relies on robust and quality industry reporting. This is why the development of the framework involves close engagement with DIGI and code signatories. Feedback provided by industry on the viability and applicability of the framework will be crucial to developing a framework that is both manageable and rigorous.

The first version of the draft measurement framework is narrow in scope, focusing on Objectives 1 and 3 of the code[6] and directed towards code signatories. This is due to the challenges associated with, and the novel nature of, this undertaking. As the framework is refined and tested as a useful insight into the effectiveness of platforms' systems and processes to address misinformation, we expect that it will be extended to major non-signatories to the code.

---

[6] Appendix A outlines in detail the outcomes of the code and signatories' commitments.

The ACMA continues to assess code signatories' transparency reports to provide additional transparency to the Australian public

Under the code, signatories with services exceeding one million monthly active Australian end users are required to provide annual transparency reports which set out progress towards relevant outcomes contained in the code. These are made available on the DIGI website. Signatories that provide a service in Australia with a user base of less than one million monthly active Australian end users are required to provide an initial report to DIGI and must update this report annually if there are any material changes to the report. However, monthly active user reporting is not provided under the code, making it difficult to determine which signatories are subject to the different reporting requirements.

The ACMA has assessed the fourth set of transparency reports. Our detailed analysis is included in the next part of this report.

Assessment of platform reporting activities

As the code is outcomes-based, it deliberately provides signatories with flexibility to implement and use measures that are proportionate to the risk of potential harm and the nature of their services. The ACMA has consistently stressed that to be both effective and enforceable, outcomes-based regulatory models must be supported by robust measurement and performance reporting frameworks.[7] In particular, we have stressed the need for KPIs which are connected to code commitments and measurable through high-quality sets of data, in particular, Australian-specific data.

We have assessed each signatory individually and developed report cards which identify the strengths and weaknesses of their transparency reports (Appendix C). We have also undertaken a detailed data analysis exercise from each transparency report published since May 2021. This forms part of our work to build a measurement framework and to look at reporting and existing data across the life of the code. We also consider this analysis timely with the receding impact of COVID-19 on misinformation, as new trends in data are likely to emerge. This year's report includes additional insight following this work.

Transparency reports continue to provide some qualitative insights into signatories' responses to misinformation

Australian-specific insights through case studies
It was positive to see signatories continue to use case studies to highlight their interventions for topics of particular social significance. Half of the signatories in this reporting period (Apple, Google, Meta, Microsoft and TikTok) provided data and qualitative commentary on the changes to their systems and processes, or new measures they implemented, ahead of the 2023 Voice Referendum in Australia. For instance, Google provided details on changes it made to its Political Content Policy and reported data about advertising spending during the Voice Referendum and the 2023 New South Wales state election.

Similarly, Meta provided details of the various measures it implemented ahead of the 2023 Voice Referendum campaign period, which included rolling out civic products such as referendum day reminders and stickers on Facebook and Instagram (which Meta reports reached 12.7 million Australian users on Facebook and 6.55 million Australian users on Instagram). We consider the inclusion of data points like these to be particularly relevant, as they provide insights into the effectiveness of Meta's interventions.

How signatories are responding to changes in technology
Some signatories addressed feedback from the ACMA's July 2023 report to take further steps to adapt quickly to technology and service changes by reporting on how they have addressed AI technologies. We note that reporting on AI technologies was an update to the Best Practice Transparency Reporting Guidelines, prepared by the code's independent reviewer and published in DIGI's 2024 annual report. These were provided to signatories to inform their reports.

---

[7] The ACMA's position paper to guide the code's development and our June 2021 and June 2023 reports to government on the code are available on our website.

For example, Google has provided information against Outcome 1a on actions taken to combat Gen AI misinformation, which include rolling out an 'About this image' feature on Google Search that gives users a way to check the credibility and context of images, as well as publishing YouTube's approach to responsible AI innovation, which requires creators to disclose when they have used AI tools to create synthetic content. Similarly, TikTok and Meta have provided information under Outcome 1a on the introduction of new labels for Gen AI content, which makes it clear to users when content is generated or significantly altered by AI technology. Meta and TikTok did not identify data points against the performance of these features.

More work needs to be done to improve quantitative reporting

Reporting against consistent Australian-specific data points
In some cases, signatories have provided consistent year-on-year Australian-specific data points under code outcomes, allowing for comparison over time. Some signatories provided trended data in their reports, with associated contextual analysis to explain changes in their reports. For example, Microsoft has provided consistent year-on-year data for Microsoft Advertising, linking the significant increase in ad rejections to network-wide policies being implemented prior to the start of the reporting period for 2023. It also noted that increases in this data and related figures may be due to its expansion into new international markets and the growth in adoption of certain advertising formats. The following figures set out the trended data collected by the ACMA from Microsoft's reports.
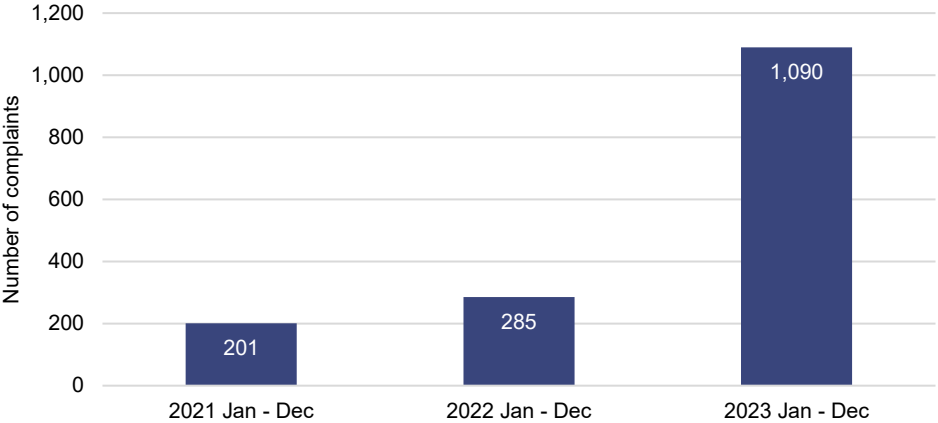
Figure 1: Australian rejections and appeals of ads[8] for containing or leading to disinformation
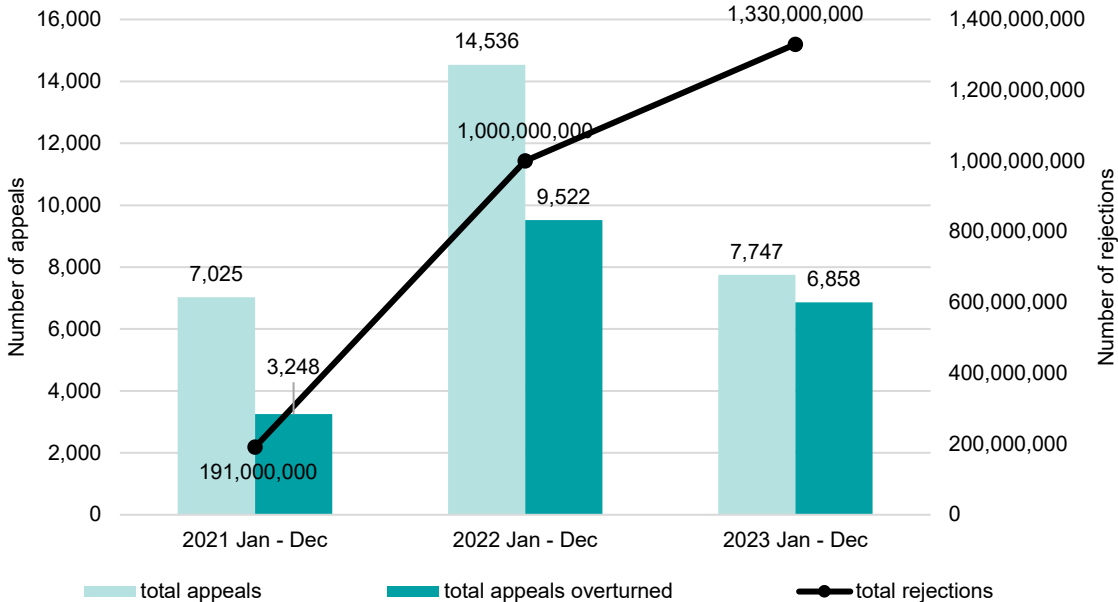
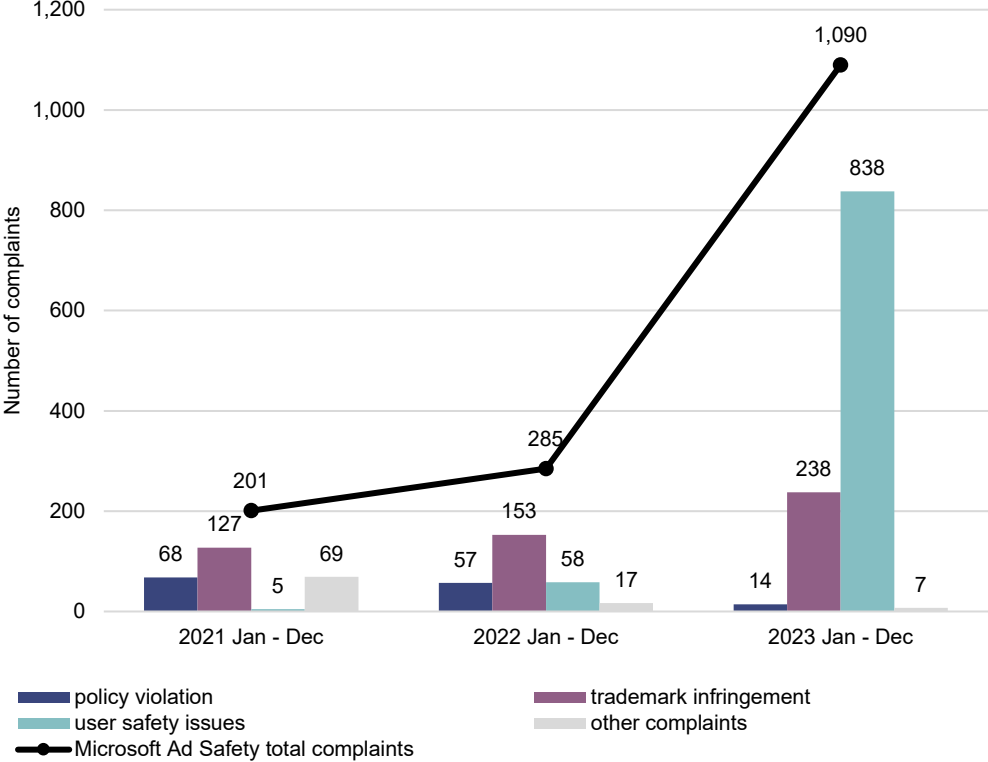*Note: ACMA aggregated graph based on Microsoft Transparency report data*

---

[8] The ACMA understands that 'total appeals overturned' refers to appeals of ad rejections, were Microsoft upholds its original decision.

Figure 2: Australian complaints about ads for containing or leading to disinformation

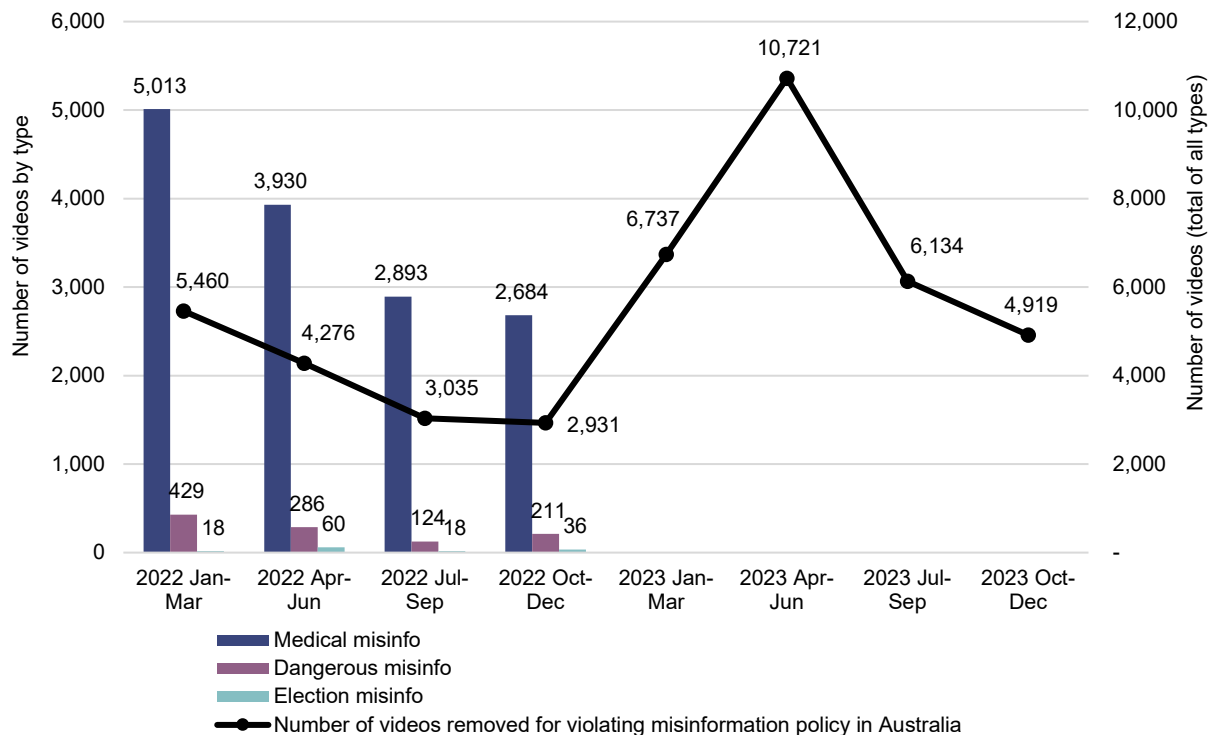*Note: ACMA aggregated graph based on Microsoft Transparency report data*

Figure 3: Types of Australian complaints about ads for containing or leading to disinformation

Note: ACMA aggregated graph based on Microsoft Transparency report data

In other cases, signatories have reported some consistent year-on-year data but have not provided the same level of granularity. For example, for the 2022 calendar year, TikTok provided breakdowns of the content removed for breaking its misinformation policies for being medical, election and dangerous misinformation policies, but did not provide data against these sub-categories in its most recent report.

Figure 4: Number of TikTok videos removed violating misinformation policy in Australia

Note: ACMA aggregated graph based on TikTok Transparency report data

One signatory did not provide any year-on-year analysis, with others only providing analysis for a limited number of data points. DIGI's independent reviewer observed that:

> It is of considerable concern when a major Signatory continues to supply only isolated annual data points. The reasoning for not providing multi-year data – that it would require contextual commentary – contradicts the core purpose of the reporting regime. This reluctance to provide accessible data, which in any case is publicly available in the previous reports, also contradicts the recommendations of the Australian Communications and Media Authority (ACMA) in its second report to government (July 2023).[9]
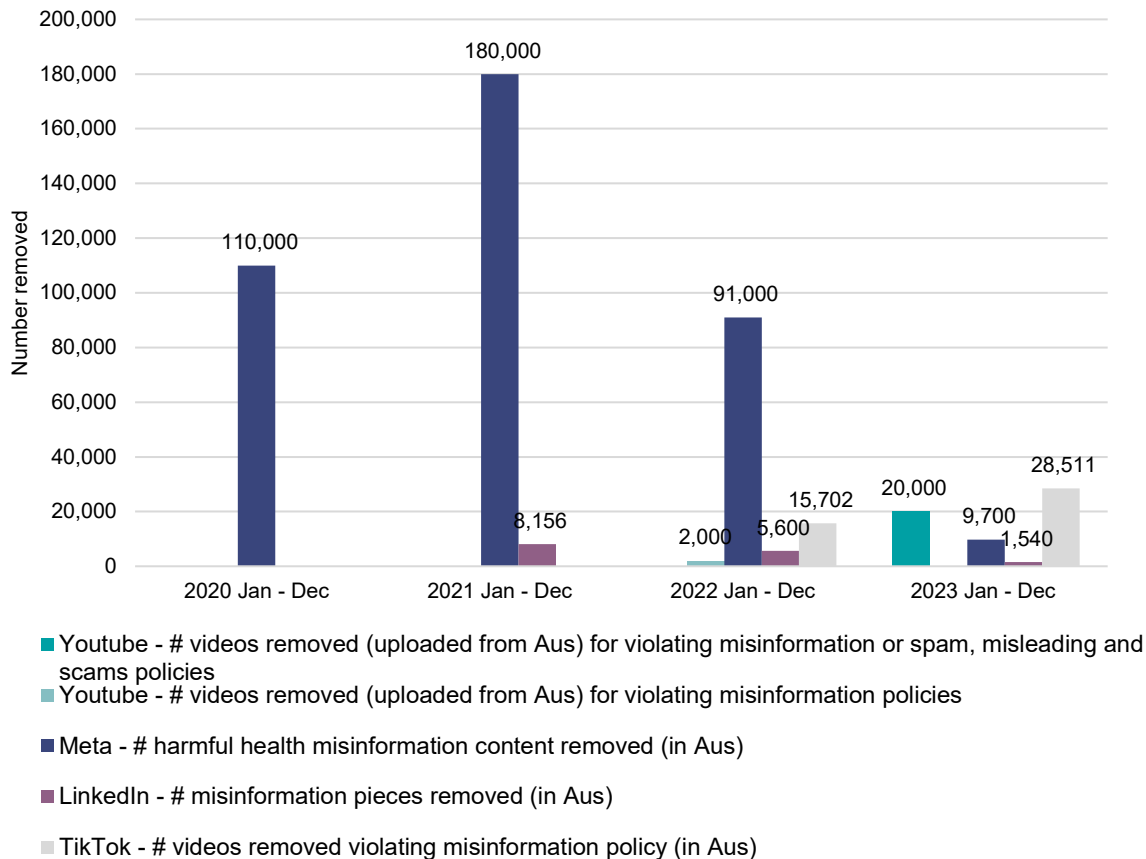
While the ACMA is collecting and creating trended data itself, the provision of contextual commentary to explain changes in data is a key function of transparency reports. This is because transparency reports are intended to demonstrate how well signatories are tracking against code outcomes. Without qualitative information to explain increases, decreases or consistent data points (and in the absence of KPIs), it is difficult to interpret the data provided or to assess how a signatory is performing in meeting its obligations.

---

[9] DIGI, Australian Code of Practice on Disinformation and Misinformation - Annual Report [PDF], DIGI website, 2024, accessed 19 June 2024, p. 9.

Limited data reported consistently across signatories

There is little, if any, consistent data across signatories. This makes it difficult to assess the impact of signatories' measures in Australia. Even where signatories report similar data points, they are difficult to compare as some signatories have not reported the same data points year-on-year and others have not reported data in every transparency report. Figure 5 sets out an example of data reported across signatories against content removal policies, although with some variation.

Figure 5: Comparison of content removal by digital platforms in Australia

Youtube - # videos removed (uploaded from Aus) for violating misinformation or spam, misleading and scams policies

Youtube - # videos removed (uploaded from Aus) for violating misinformation policies

Meta - # harmful health misinformation content removed (in Aus)

LinkedIn - # misinformation pieces removed (in Aus)

TikTok - # videos removed violating misinformation policy (in Aus)

Note: ACMA aggregated graph based on Google (YouTube), Meta, Microsoft (LinkedIn) and TikTok Transparency report data

Unlike other signatories, Meta appears to report data under its Harmful Health Misinformation Policy only, rather than all misinformation policies. It is unclear from Facebook's transparency report for the 2020 year whether the data is for Facebook and Instagram or Facebook only. For the 2022 calendar year, Google (YouTube) reported data about its content removals under its Misinformation Policy. However, for the 2023 calendar year, Google (YouTube) has reported a single combined metric covering content removals under both its Misinformation and Spam, Deceptive Practices and Scams policies (which the ACMA understands remain distinct policies). This appears to conflate the data and makes it harder to interpret the volume of content removals for misinformation. Google did not include commentary to explain the change in reporting.

As the code enters its fourth year, we expect signatories to report consistent year-on-year data with clear explanations about changes in data and reporting approaches.

Identifying new Australian-specific data points

Signatories continue to make incremental progress towards identifying additional Australian-specific data points to measure the effectiveness of their systems and processes. The following table summarises additional data points identified in this year's transparency reports. It is important that, where new data points are identified, signatories continue to report against them or provide information to explain why they have been discontinued.

Table 1: Examples of new Australian-specific data points reported by signatories[10]

Signatory	New Australian-specific data points
Google	Google reported 1 new data point under Outcome 1a for YouTube in relation to the percentage of videos (with 10 or fewer views) removed from Australian IP addresses for violating community guidelines. Google reported 6 new data points under Outcome 1c[11] for YouTube and AdSense. The new data points relate to appeals (for example, number of appeals received in Australia about YouTube Community Guidelines violation removal decisions). Google identified 3 new data points under Outcome 4 which related to Google News Lab's distribution of fact checks to Australian news publishers, Google-supported placements of Indigenous trainees in newsrooms, as well as YouTube's media literacy campaign 'Hit Pause'. Google reported 11 new data points under Outcome 5 which related to ad spending for the NSW state election and the Voice Referendum (broken down by the top 3 advertisers, states/territories, and the format of ads). Google identified 2 new data points under Outcome 6 which related to the Newshound media literacy program for students.
Meta	Meta reported 2 new data points under Outcome 2 which provide data on the number of ads removed for violating Misinformation policies in Australia, and the number of ads removed for not complying with Meta's Social Issues, Elections and Politics ad policies in Australia. Meta identified 4 new data points under Outcome 4 which relate to engagements of the civic products implemented on Facebook and Instagram for the Voice Referendum. Meta reported 2 new data points under Outcome 6 which relate to engagements of the media literacy campaign on the Voice Referendum on Facebook and Instagram.

---

[10] The ACMA's data analysis exercise has also included an assessment of which outcome the data points best reflect. We note that signatories do not always report data against the most relevant outcome.

[11] In Google's transparency report, these six data points are not categorised under Outcome 1c; they are instead under Objective 1 (p.8) and Objective 2 (p.16). However, the ACMA considers them to be most applicable to Outcome 1c.

Microsoft	Microsoft identified 3 new data points under Outcome 1a which relate to the number of accounts opted in, as well as successfully verified, for the Advertiser Identity Verification (AIV) program, and the number of Microsoft Start comments proactively blocked in Australia. Microsoft reported 1 new data points under Outcome 4 on its engagement with NewsGuard, which reports ratings for news and information sites in Australia and New Zealand.
TikTok	TikTok reported 4 new data points under Outcome 1a which relate to video removals for violation of TikTok's community guidelines. TikTok reported 7 new data points under Outcome 4 on engagement of notice tags and resources in the Voice Referendum Hub and Voice Referendum Search Guide, as well as number of videos fact-checked during the Voice Referendum (including those escalated and enforced). TikTok identified 1 new data point under Outcome 6 on the number of partners introduced to the Community Partner Channel in Australia.

However, many signatories still have not identified any data points against code commitments, making their progress towards meeting these outcomes difficult to track. The following table identifies outcomes signatories have committed to but have not identified any relevant metrics for. We note that this was Twitch's first transparency report.

Table 2: Signatories' outcome commitments with no data points reported against

Signatory	Outcome commitment with no identified Australian-specific metric[12]
Adobe	Outcome 1a, 4, 6
Apple	Outcomes 1e, 2, 6
Google[13]	Outcomes 1b, 1d, 1e, 3
Meta[14]	Outcomes 1b, 1c, 1e, 5
Microsoft[15]	Outcomes 1b, 1e, 2, 5, 6
Redbubble	Outcomes 1a, 1b, 1c, 1d, 1e, 2, 3, 4, 6
TikTok[16]	Outcomes 1b, 1c, 1e, 2, 3, 5
Twitch	Outcomes 1a, 1b, 1c, 1d, 1e, 2, 4, 6

---

[12] Appendix A outlines in detail the outcomes of the code and signatories' commitments.
[13] While Google did not identify any specific data points against Outcome 3, the ACMA considers 1 data point identified against Outcome 1a, is also applicable.
[14] While Meta did not identify any specific data points against Outcome 1d and 3, the ACMA considers some data points identified against Outcome 1a are applicable to these outcomes.
[15] While Microsoft did not identify any specific data points against Outcome 1c, the ACMA considers some data points identified against Outcome 1a are applicable to this outcome.
[16] While TikTok did not identify any specific data points against Outcome 1d, the ACMA considers some data points identified against Outcome 1a are applicable to this outcome.

Data integrity issues limit insights from quantitative data reported

While signatories have put in significant effort to collect and report data annually, across the board, there does not appear be a consistent approach to reporting data year on year. Although the Best Practice Transparency Reporting Guidelines, developed by the code's independent reviewer supports better reporting, data validation and quality assurance appear to be outside the scope of this role and these guidelines.

Considerable effort and time have been required to collect and collate data reported by signatories. For much of the data, the ACMA has had to cross reference data points within transparency reports and between years, as well as to use contextual commentary to identify whether some data points have been reported consistently or are Australian-specific. For example, in its transparency report for the 2022 calendar year, TikTok provided data about its content removal rates for harmful misinformation. The ACMA was only able to confirm these 2022 reported numbers were Australian-specific by comparing the numbers against a graph included in its transparency report for the 2023 calendar year.

In other cases, there remains some ambiguity in the metrics reported. For example, in previous reports, Meta reported data for content removals under its Harmful Health Misinformation Policy. In this year's report it refers to misinformation policies. However, contextual commentary (which explains a drop off in removals as a result of updates to its policies stemming from the changing impact of COVID-19) and year-on-year data (which matches data previously reported under its Harmful Health Misinformation Policy), suggest that the 2023 calendar year data also relates to its Harmful Health Misinformation Policy specifically, rather than misinformation policies generally. We also note that in its most recent transparency report, Meta referred to a past data point of 180,000 pieces of content removed globally in 2021 for violations of its misinformation policies. However, in its transparency report for the 2021 reporting year, it stated that 180,000 pieces of content were removed from accounts specific to Australia.

The inconsistent way data is reported, may also account for seemingly inaccurate data. For example, in TikTok's most recent transparency report, trended data reported against its Integrity and Authenticity policies for 2022 is identified as Australian.[17] However, a comparison with its transparency report for the 2022 calendar year and its Community Guidelines Enforcement Report show that this 2022 data is most likely global data.[18] Microsoft also appears to have reported some inaccurate data, with total complaints figures not matching complaint breakdowns.

In our view, transparency reports should allow interested members of the public, academics, researchers and government agencies to easily access and extract information. The significant effort required to extract and verify data significantly limits the utility of information reported by signatories. While each of the examples outlined above may not have been the result of any deliberate effort to obscure or confuse data reported,[19] these data integrity issues are indicative of broader systemic issues with how data is reported. A more robust reporting framework would help prevent ambiguity and assist signatories to avoid potential mistakes.

---

[17] See Figure 2B TikTik 2023 Annual Transparency report, p. 8.
[18] TikTok 2022 Annual Transparency report, p. 19; See also "Total video removal, by policy", Community Guidelines Enforcement Report, accessed 30 July 2024.
[19] On 24 July 2024, Ofcom fined TikTok £1.875 million for failing to accurately respond to a data request. Ofcom's investigation, among other matters, found a number of failings in TikTok's data governance and provided inaccurate data in a submission.

For the code to be successful, the ACMA and the public must be able to use the information reported and have confidence that the information provided is reliable. In our view, this year's transparency reports reinforce the need for industry to work with the ACMA to develop, adopt and implement a measurement framework.

Our data extraction and analysis remain ongoing. We expect to work with signatories to overcome these data integrity issues as we implement the framework. We hope to provide a thorough analysis of all reporting years in 2025.

Data points tend to focus on content removal
The code encourages signatories to adopt proportionate measures to address the harms posed by misinformation, considering a range of factors, such as:

- the actors which are engaged in propagating misinformation
- the proximity and severity of the harm that is reasonably likely to result from the propagation of the content
- the need to protect freedom of expression in balance with other human rights.

In some cases, the most appropriate form of action may be content removal. Google (YouTube)[20], Meta (Facebook[21], Instagram[22]), Microsoft (LinkedIn)[23], Redbubble[24], TikTok[25] and Twitch[26] all have policies to remove certain types of misinformation. These policies share a similar threshold of removing harmful misinformation including health and public safety misinformation, and civic and electoral misinformation.

Signatories adjust these policies based on an assessment of harms. As noted above, Meta accounted for the reduction in misinformation content removals during the 2023 calendar year because of changes to its Harmful Health Misinformation Policy. Specifically, Meta identified a change in attitudes towards the potential impact of COVID-19 misinformation.

In many cases, balancing the risk of harms from misinformation with other human rights, such as freedom of expression, necessitates more nuanced system and policy responses from code signatories. We encourage this nuanced and balanced approach. Google (YouTube)[27], Meta (Facebook[28], Instagram[29]), Microsoft (LinkedIn[30]), and TikTok[31] all have policies to manage certain types of misinformation but not remove it. For instance, YouTube may allow content that violates its misinformation policies if it includes additional context in the video, audio, title or description for educational, documentary, scientific or artistic purposes. Similarly, Facebook, Instagram, LinkedIn and TikTok all have policies to apply warning labels to misinformation content and/or to restrict its reach. Restricting the reach of misinformation is an algorithmic response and may result in downranking content in feeds or

[20] YouTube, Misinformation policies - YouTube Help (google.com), YouTube Help website, 2024, accessed 12 July 2024.
[21] Facebook, Misinformation | Transparency Centre (meta.com), Facebook Community Standards website, 2024, accessed 12 July 2024.
[22] Instagram, How Instagram addresses false info | Instagram Help Centre (facebook.com), Instagram Help Centre website, 2024, accessed 12 July 2024.
[23] LinkedIn, Misinformation and inauthentic behavior | LinkedIn Help, LinkedIn Help website, 2023, accessed 12 July 2024.
[24] Redbubble, Community and Content Guidelines – Redbubble, Redbubble Help Centre website, 2023, accessed 12 July 2024.
[25] TikTok, Combatting Harmful Misinformation, TikTok Transparency Centre website, 2024, accessed 12 July 2024.
[26] Twitch, Preventing Harmful Misinformation Actors on Twitch, Twitch Safety website, 2022, accessed 12 July 2024.
[27] YouTube, Misinformation policies - YouTube Help (google.com), YouTube Help website, 2024, accessed 12 July 2024.
[28] Facebook, Fact-checked misinformation | Transparency Centre (meta.com), Facebook Transparency Centre website, 2024, accessed 12 July 2024.
[29] Instagram, Reducing the spread of false information on Instagram, Instagram Help Centre website, 2024, accessed 12 July 2024.
[30] LinkedIn, False or misleading content | LinkedIn Help, LinkedIn Help website, 2023, accessed 12 July 2024.
[31] TikTok, Combatting Harmful Misinformation, TikTok Transparency Centre website, 2024, accessed 12 July 2024.

minimising the speed or ease by which such content can be reshared. Some platforms may use a combination of these measures.

Data points identified by signatories in their transparency reports tend to focus on content removal, with few data points identified for other types of action taken to manage misinformation, but not remove it. Measuring the effectiveness of these types of actions, for example the effectiveness of warning labels by providing data about Australian end-user engagement with content which contravenes policies to manage misinformation, is necessary to provide sufficient transparency on the full suite of measures signatories are taking to address misinformation content.

As noted above, Meta provided some metrics about the reach and impressions of its civic products during the Voice Referendum, see Figure 6 below. While these metrics provide meaningful insights, a fuller picture of Meta's interventions could be drawn if data were also provided on the reach and impressions on content identified by Meta as misinformation during the same period, such as the reach and impressions on content it labelled with a warning label. Figure 7 sets out trended data Meta reported for content applied with a warning label based on debunking articles written by fact checking partners in Australia for the whole of 2022 and 2023.

Figure 6: Trends for the Voice Referendum in Australia 2023 – Facebook and Instagram
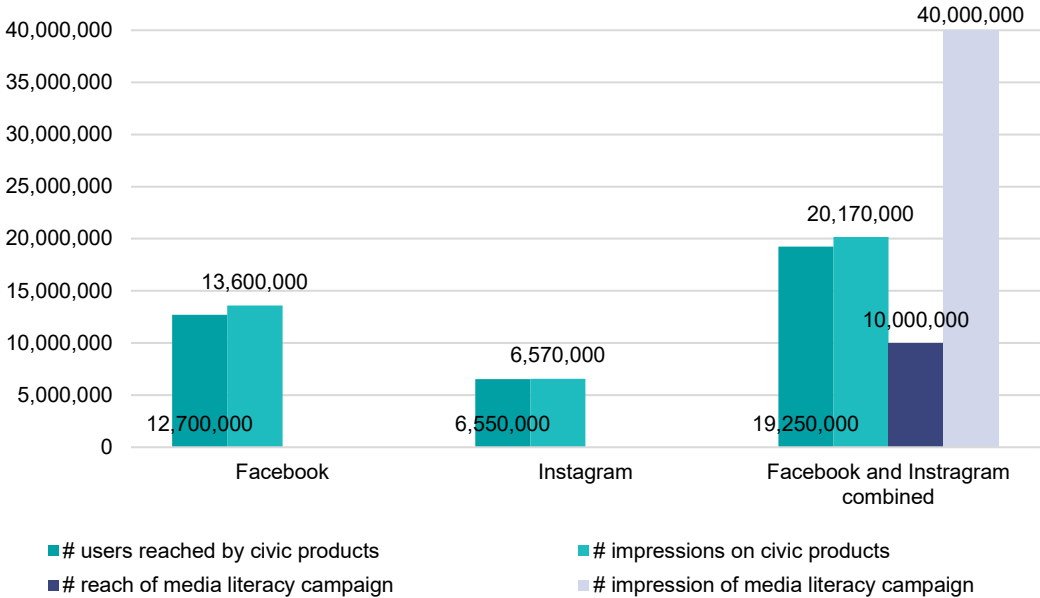
Note: ACMA aggregated graph based on Meta transparency report data

Figure 7: Number of distinct pieces of content (including reshares) with warning label in Australia, based on debunking articles written by fact checking partners in Australia – Facebook and Instagram
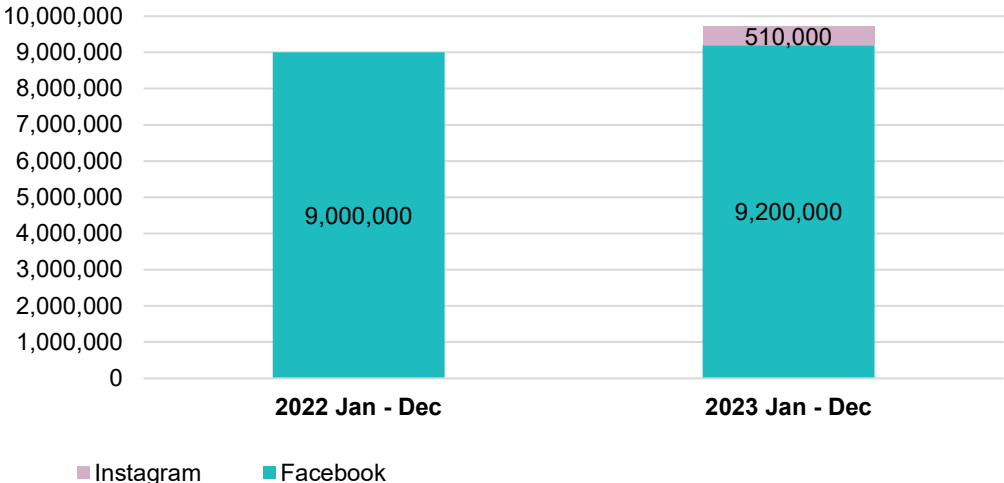
Note: ACMA aggregated graph based on Meta Transparency report data

No discernible progress towards identifying key performance indicators

We have not identified any discernible progress by signatories towards identifying KPIs. This was also noted by DIGI's independent reviewer, who has observed that:

> It has been a recurring recommendation, both of ACMA and this auditor, that Signatories formally commit to internally consistent Key Performance Indicators (KPIs). The fact this has not happened is an indicator that, as stated, improvement of the reporting regime has stalled.[32]

While signatories have reported Australian-specific data, and in some cases reported year-on-year data, without KPIs it is difficult to determine whether signatories consider their efforts to be successful or to hold them accountable to their commitments.

As noted above, to address some of these gaps, the ACMA has been progressing work to develop a framework to measure the effectiveness of signatories' measures at meeting the code's objectives. Appendix B sets out further detail on how this framework is intended to operate, including that signatories develop their own internal KPIs against industry wide metrics.

While acknowledging the challenging task of developing industry-wide metrics, it is disappointing that signatories have not made progress towards identifying internal KPIs for data they already report year on year.

---

[32] DIGI, Australian Code of Practice on Disinformation and Misinformation - Annual Report [PDF], DIGI website, 2024, accessed 19 June 2024, p. 9.

We note Meta's Oversight Board recently articulated the importance of developing KPIs to assess the effectiveness of Meta's voting interference rules under its Coordinating Harm and Promoting Crime Policy. The Oversight Board stated that it:

> acknowledges Meta's efforts on the Voice Referendum. The company proactively identified potentially violating content under the voting interference rules of the Coordinating Harm and Promoting Crime and Misinformation Community Standards. The phrases "double vote" and "vote multiple times" were the keywords that activated the company's keyword-based detection system in this case. According to Meta, the system is adapted to local contexts. Based on the information shared, the Board notes that initiatives like these should be consistently applied across the globe, in countries undergoing elections, although Meta is encouraged to develop success metrics for assessing how effective keyword-based detection is.[33]

Concluding remarks

It is positive that signatories have continued to provide some meaningful insights into their systems and processes to address misinformation on their platforms, including via case studies. However, we are concerned by the minimal steps taken to significantly improve the quality and usability of data reported and it is disappointing that signatories have made no discernible progress towards developing KPIs.

Additionally, interpreting data reported in transparency reports remains challenging. We consider that observations in our July 2023 report about signatories' reporting still urgently need to be addressed. We note that the code's independent reviewer also observed the lack of progress towards KPIs in this year's transparency reports and that data points continue to be reported in isolation by some signatories.

In our view, this most recent set of transparency reports reinforces the need for signatories to proactively adopt and implement a more comprehensive reporting regime to ensure the continued success of voluntary arrangements.

---

[33] Oversight Board, Oversight Board Upholds Meta's Decisions in Australian Electoral Commission Voting Rules Cases | Oversight Board, Oversight Board website, 2024, accessed 11 July 2024.

Governance and administration

Robust governance and administration arrangements will provide greater reassurance that voluntary arrangements are sustainable

Our consistent messaging to industry has been that robust reporting and governance frameworks are critical to the success of the code. In our July 2023 report, we noted that the absence of these elements (including the complaints mechanism) from the scope of the 2022 review was a missed opportunity to assess whether the frameworks are functioning effectively.

In its 2024 annual report, DIGI has stated that signatories have decided to defer a further review of the code.[34] The reason being that the status of the draft Bill is unclear and it would be desirable to conduct the review in a state of greater regulatory certainty.

While we consider this approach reasonable under the circumstances, the delay increases the need for, and our expectation of, a thorough consideration of the code's reporting and governance frameworks during the next review. Incremental improvements will not be sufficient to provide reassurance that voluntary arrangements can provide adequate community safeguards and that further regulatory intervention, in the form of co-regulatory codes or industry standards, is not required.

Delaying the code review means that industry should also have sufficient time to consider specific matters raised in this report and in our July 2023 report about:

- the scope of the code (which have not been readdressed in this report)
- the code's reporting framework
- the governance and administration of the code.

We also continue to encourage greater participation in the code, particularly from platforms with large active Australian user bases. Each new signatory contributes to the objective of the code to provide a minimum level of protection for Australian users of digital platforms. It has been positive to see DIGI's efforts in this respect and we hope to continue to work with DIGI to expand the number of code signatories.

Code governance

In our July 2023 report, we commented that information provided in DIGI's annual reports focused on the existence and roles of oversight mechanisms but did not provide much visibility about the operation of these mechanisms. DIGI has included some additional information about the activities of its Administration Sub-committee during 2023 in its most recent annual report. Details include the number of times it met, the matters it considered and updates to its processes, such as that signatories will provide the committee with updates concerning material changes in the functionality of their service and relevant policy and service updates. We encourage continued improvements to transparency as the code continues to mature.

---

[34] Clause 7.8 of the code sets out that the code will be reviewed at 2 yearly intervals.

Independent review process for transparency reports

In its annual report, DIGI states that in March 2024 it asked its independent reviewer to update the Best Practice Transparency Reporting Guidelines to:

- include specific requests for clear explanations of major changes in policy information
- include specific information about efforts to combat Gen AI misinformation
- update the table of code signatories to reflect withdrawals and additions
- highlight areas for continual improvement based on analysis of past reports including KPIs.

The report states that these changes were made in response to the expansion of the code following the last review and in response to recommendations made by the ACMA in our July 2023 report. We understand that this guideline was provided to signatories prior to drafting the latest transparency reports. As discussed above, signatories have provided information about their responses to AI in response to the revised guidelines.

We consider these to be positive additions. We would welcome the opportunity to engage directly with DIGI and the code's independent reviewer on the next iteration of the guidelines as we progress towards implementing a measurement framework. Clarifying the relationship between transparency reports and a measurement framework will be especially important to assist signatories to prepare for their next reports.

Complaints facility and clarity on code comments

In its annual report, DIGI has provided some details about a social media advertising campaign to raise awareness of the code among Australians.[35] DIGI has also commented on the 2 Reset Australia complaints about X Corp and Meta. Based on the figures listed, these were the only complaints considered by DIGI's independent complaints sub-committee. Overall, complaint numbers seem low considering community concern about misinformation.

Signatories with relevant services, except for Twitch and TikTok, did provide some quantitative data about reports, complaints and appeals captured under Objective 1 of the code in the recent transparency reports. It is intended that the implementation of the ACMA's measurement framework will support this reporting by introducing a common set of metrics.

On 8 December 2023, government requested industry to develop a voluntary internal dispute resolution standard by July this year.[36] In our view, the code complaints facility under the code should nonetheless be strengthened, including to establish a clearer referral pathway between signatories and DIGI's complaints facility (where escalation is appropriate), in conjunction with increased transparency on signatory commitments, to be considered in the next code review.

Relatedly, we recommend that the DIGI website provides greater clarity on the outcomes signatories have adopted to meet the objectives they have opted into under the code. While implementation of the measurement framework will provide more clarity to signatories' commitments, reporting against the framework is expected to occur annually with transparency reports. Greater clarity is necessary for the effective functioning of a complaints facility, as members of the public should be able to easily reference signatories' commitments to help inform their decision to make a complaint.

---

[35] DIGI, Australian Code of Practice on Disinformation and Misinformation - Annual Report [PDF], DIGI website, 2024, accessed 19 June 2024, p. 13.

[36] Australian Government, Government's response to the ACCC Digital Platform Services Inquiry, Treasury website, 8 December 2023, accessed 28 June 2024.

Next steps

On 1 July 2024, the ACMA announced its [compliance priorities](#) for 2024–25.

Combating misinformation and disinformation on digital platforms

Misinformation can pose a risk to the health and safety of individuals, as well as society more generally. When it is spread deliberately (disinformation), it causes confusion, undermines trust in governments and institutions, and can contribute to financial harm for Australians.

In 2024–25, we will continue to oversee the performance of digital platforms under the voluntary Australian Code of Practice on Disinformation and Misinformation. We are also working with the Australian Government to develop and strengthen our regulatory powers and, if new legislation is introduced, we will move to focus on implementation and addressing digital platforms' compliance with the legislation.

Starting to implement a measurement framework

As articulated in this report, we consider that industry can do more to provide transparency to the Australian public on the effectiveness of the measures they have put in place to meet their code commitments. The ACMA developed the first iteration of a measurement framework to address the issues with the existing reporting framework identified in our July 2023 report. In our view, the framework will also help to address many of the data integrity and extraction issues we have identified in this report.

We look forward to working constructively with industry to implement the framework. We expect this process will involve ongoing consultation with signatories to arrive at a workable first iteration for implementation. Ideally, the code's reporting guidelines will be updated, and signatories will report against the first iteration of the framework as part of their next transparency reports.

Signatories' transparency reports will remain an important and necessary part of the code's reporting framework. Transparency reports provide signatories the opportunity to report qualitative information. Such information may explain changes in year-on-year data or the impetus for updates and changes to their systems and policies. Transparency reports will also continue to provide signatories the opportunity to report on their efforts to address specific harms (for example, from specific events) caused by misinformation. As noted above, the most recent transparency reports continued to provide relevant and valuable information on specific events, such as the Voice Referendum and Russia's invasion of Ukraine. In previous years, signatories have provided similarly meaningful insights into their steps to address the COVID-19 pandemic. Transparency reports will also allow signatories to continue to report data points not captured by the industry-wide metrics.

The ACMA expects that the framework will evolve over time and future iterations may be published on the ACMA's website.

A robust voluntary code supported by regulatory powers

We have commenced work to prepare for the potential introduction of new legislative powers to regulate digital platforms. If legislation is passed, we consider that using information-gathering and transparency powers (including the power to publish collected information) will be the most effective way to strengthen current voluntary arrangements and improve safeguards for the Australian public. This is because transparency frequently drives accountability and will provide much needed information to both the ACMA and the Australian public, to assess the adequacy of the voluntary code.

The accompanying guidance note to the draft Bill notes that an effective self-regulatory scheme is the preferred approach to combating misinformation on digital platforms. Code registration and standard-making powers are intended as a regulatory backstop.

As discussed in the first 2 sections of this report, a key gap with existing voluntary arrangements is the lack of transparency about the effectiveness of signatories' measures to combat misinformation. In the absence of such transparency, platforms cannot be held accountable to their code commitments, nor can it be determined whether the systems and processes signatories have put in place to meet the code's outcomes are in fact providing adequate safeguards.

Record-keeping rules

Data reported by signatories against the measurement framework will provide us with insights into how a platform's systems and processes are working. Should this information not be provided voluntarily, the ACMA could make record-keeping rules which would compel signatories to collect relevant data under the framework and report it to the ACMA. Such rules could also be applied to major non-signatories to the code, to compare efforts by signatories and non-signatories alike.[37] The ACMA intends to publish information collected to demonstrate to Australian users how individual platforms are performing.

---

[37] The ACMA would consult before making record keeping rules as part of the usual consultation processes under the *Legislation Act 2003*. Record keeping rules, as contemplated by the draft Bill, would also be disallowable by Parliament.

Case study: X Corp loses its signatory status

In November last year, X Corp (formerly Twitter), an original signatory to the code, had its signatory status withdrawn. The code's independent complaints sub-committee found that X Corp had committed a serious breach of the code for failing to provide a mechanism to the public to make reports of breaches of its policies for an extended period. It also noted that X Corp refused to cooperate with DIGI or undertake any remedial action. The decision to withdraw X Corp's signatory status was the first decision of DIGI's complaints sub-committee under the code.

To date, the ACMA is not aware of a formal response from X to the sub-committee's decision, nor did X Corp respond to the ACMA's own enquiries on the same matter.

This decision both reinforces the integrity of the code, while also highlighting some of the limits of voluntary arrangements. Under a voluntary scheme, signatories cannot be compelled to cooperate with investigative processes. The loss of signatory status is a blunt instrument for handling serious non-compliance, especially given that loss of signatory status means the loss of transparency afforded through the transparency reporting process.

The ability to make record keeping rules is critical to compel major digital platforms to continue to provide transparency about their systems and processes if they are unwilling to participate in the voluntary scheme.

Information gathering powers

Individual information requests would also play a significant role in providing necessary transparency. As discussed above, the ACMA expects transparency reports to remain a pivotal aspect of reporting under voluntary arrangements. However, given these reports are made annually, they may not provide sufficiently timely information following events of significant community concern.

The ongoing need for these types of powers has recently been highlighted by platform responses to information requests the ACMA made following the April 2024 Bondi Junction and Wakeley Church stabbing attacks.

Case study: ACMA engaged with digital platforms following the April Bondi Junction and Wakeley Church attacks in April 2024

On 17 April 2024, the ACMA requested information from code signatories, as well as major non-signatories Snap Inc, X Corp, Yahoo, Reddit, Discord, and Telegram, about their responses to misinformation which circulated widely online following the Bondi Junction and Wakeley Church attacks in Sydney in April 2024.

Platforms generally responded promptly and provided constructive information. Some provided information about why the viral spreading of misinformation does not occur on their platforms. Others provided information about the systems and processes they have in place to address harms from misinformation following unprecedented events. In some cases, this was supported by data points giving an indication of their effectiveness.

While the ACMA did meet with X Corp to discuss the request, we have not to-date received the requested information. This is particularly concerning given analysis published by the Australian Broadcasting Corporation[38] which tracked the spread of misinformation on the platform following the Bondi Junction attack.

We were also disappointed that the information provided by other platforms did not, in some cases, include the requested information about the effectiveness of systems. While it appears that some misinformation following the Bondi Junction attack originated on X Corp's platform, information does not stay siloed and can spread rapidly between platforms.

It is understandable and necessary that platforms' immediate concern following such events will be to respond to violent and terrorist related content, where content removal may be both the most effective and necessary response. However, as noted above, responding to misinformation from specific events often requires ongoing efforts, including by means other than content removal. As we have seen both internationally and in Australia, such events often lead to conspiracy theories and the spread of misinformation which can have long lasting impacts.

The limited and asymmetric information provided, reflects digital platforms' unwillingness to voluntarily engage with the ACMA. This behaviour highlights the urgent need for information gathering powers. In the absence of such powers, the ACMA has been unable to provide transparency to the Australian public about the effectiveness of platform responses after these events.

---

[38] Kevin Nguyen and Michael Workman, 'Benjamin Cohen was falsely accused of the Bondi Junction stabbings. Here's how the lie spread around the world', ABC News, 15 April 2024, accessed 20 June 2024.

Under the Digital Services Act (DSA), in October 2023 the European Commission (the Commission) has sent formal requests[39] for information to X Corp[40], Meta (Facebook and Instagram)[41] and TikTok[42] to provide information about the measures they have taken to address the spread of illegal content and disinformation on their platforms relating to the Israel-Hamas conflict.[43] Under the DSA, the Commission can send requests for information to verify a platform's compliance. Since coming into force, it has sent a broad range of requests for information such as:

- formal requests to 17 digital platforms, including Facebook, Google Search, Instagram, LinkedIn, Snapchat, TikTok and YouTube, to provide more information on their compliance with researcher data access obligations.

- formal requests to Bing and Google Search, Facebook, Instagram, Snapchat, TikTok, YouTube, and X to provide more information on their respective mitigation measures for risks linked to Gen AI, such as so-called 'hallucinations' where AI provides false information, the viral dissemination of deepfakes, as well as the automated manipulation of services that can mislead voters.

- other requests for information to Apple, Amazon, Google, Meta (x2), Snap, TikTok, YouTube and X to provide more information on measures they have taken to comply with various obligations related to their risk assessments and mitigation.

---

[39] European Commission, DSA: Fighting disinformation (europa.eu), EU Commission website, 18 October 2023, accessed 15 July 2024.

[40] European Commission, Commission sends request for information to X under DSA, EU Commission website, 12 October 2023, accessed 15 July 2024.

[41] European Commission, Commission sends request for information to Meta under the Digital Services Act, EU Commission website, 19 October 2023, accessed 15 July 2024.

[42] European Commission, Commission sends request for information to TikTok under the Digital Services Act, EU Commission website, 19 October 2023, accessed 15 July 2024.

[43] Under the DSA, the Commission has formal investigative and enforcement powers to monitor the compliance of very large online platforms and search engines (VLOP/VLOSEs) with their obligations under the DSA.

Conclusion

We consider that the substantive issues raised in our previous reports remain in urgent need of attention if voluntary arrangements are to be successful. The data integrity issues identified in our analysis of data reported by signatories, combined with the difficulties faced by the ACMA in obtaining information from some major digital platforms about matters of significant community concern, reinforces the need for formal regulatory powers, in particular, information-gathering transparency powers, to hold platforms accountable.

The ACMA will continue to work with industry to implement the first iteration of a measurement framework. We will also support DIGI and the code's independent reviewer to further develop the Best Practice Transparency Reporting Guidelines to support the framework's introduction.

In 2025, we also expect industry to undertake a review of the code and anticipate making a further submission to this process. Robust reporting and governance arrangements are critical to the success of the code, and proactive, voluntary adoption of such arrangements will provide reassurance that voluntary arrangement can continue to be successful.

As noted above, should the Bill be passed by parliament, it is our view that supporting and improving voluntary arrangements in the first instance will be critical to providing immediate and ongoing safeguards to Australians from misinformation. The combination of voluntary reporting by industry and regulatory levers to increase transparency across the broader industry will also provide much needed transparency on the impact of the code.

Appendix A: List of signatories' commitments under the code

Outcomes	Services[44]							
	Meta (includes Facebook and Instagram)	Google (includes Search, News, Ads and YouTube)	TikTok	Apple (includes Apple News)	Adobe	Redbubble	Microsoft (includes Bing, Start Advertising, LinkedIn)	Twitch
Mandatory commitment - 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.	✓	✓ (S, N, A, Y)	✓	✓	✓	✓	✓ (B, S, A L)	✓
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.	✓	✓ (S, N, A, Y)	✓	n/a	n/a	✓	✓ (S, A, L)	✓
1c: **Users can report** content and behaviours to Signatories that violates their policies under 5.10 through publicly available and accessible **reporting tools**.	✓	✓ (S, N, A, Y)	✓	✓	n/a	✓	✓ (B, S, A, L)	✓
1d: Users will be able to access **general information about Signatories' actions** in response to reports made under 5.11.	✓	✓ (S, N, A, Y)	✓	n/a	n/a	✓	✓ (B, S, A, L)	✓

---

[44] As a signatory to the code, Legitimate will provide its first transparency report in May 2025, outlining their commitments to the Code and their approach to misinformation and disinformation.

| Commitment								
1e: Users will be able to access g**eneral information about Signatories' use of recommender systems** and have options relating to content suggested by recommender systems.	✓	✓ (S, N, A, Y)	✓	✓	n/a	✓	✓ (L)	✓
2: **Advertising** and/or monetisation incentives for Disinformation are reduced.	✓	✓ (A, Y)	✓	✓	n/a	✓	✓ (A, L)	✓
3: The risk that Inauthentic User Behaviours undermine the **integrity** and **security** of services and products is reduced.	✓	✓ (S, N, A, Y)	✓	n/a	✓	✓	✓ (S, A, L)	✓
4: Users are enabled to make more informed choices about the source of news and factual content accessed via digital platforms and are better equipped to identify Misinformation.	✓	✓ (S, A, Y)	✓	✓	✓	✓	✓ (B, S, L)	✓
5: Users are better informed about the source of **Political Advertising.**	✓	✓ (A)	✓	n/a	n/a	n/a	✓ (A, L)	✓
6: Signatories support the efforts of independent **researchers** to improve public understanding of Disinformation and Misinformation.	✓	✓ (S, Y)	✓	✓	✓	✓	✓ (Microsoft)	✓
Mandatory commitment -7: The **public can access information about the measures** Signatories have taken to combat Disinformation and Misinformation.	✓	✓ (S, A, Y)	✓	✓	✓	✓	✓ (B, S, A, L)	✓

Appendix B: The draft Measurement Framework

The first iteration of the framework will focus on:

- Objective 1 of the code, to provide safeguards against harms that may arise from misinformation This objective is the key mandatory operative objective to which all signatories have agreed

- Objective 3 of the code, to work to ensure the integrity and security of services and products delivered by digital platforms. While not mandatory, this offers important insights into measures directed at addressing coordinated inauthentic behaviour.

The framework adopts an intervention-agnostic approach. This approach recognises the suite of measures signatories may use to address misinformation content and this is a key focus of the measures and metrics in the measurement framework. Despite this intervention-agnostic approach, we are keen for the final framework to shed more light on the signatories' content removal rates as well as complaints and appeals processes, noting general community concern about freedom of expression.

Figure 8 sets out the structure of the framework, which is divided into 3 parts.

Figure 8: Measurement Framework structure
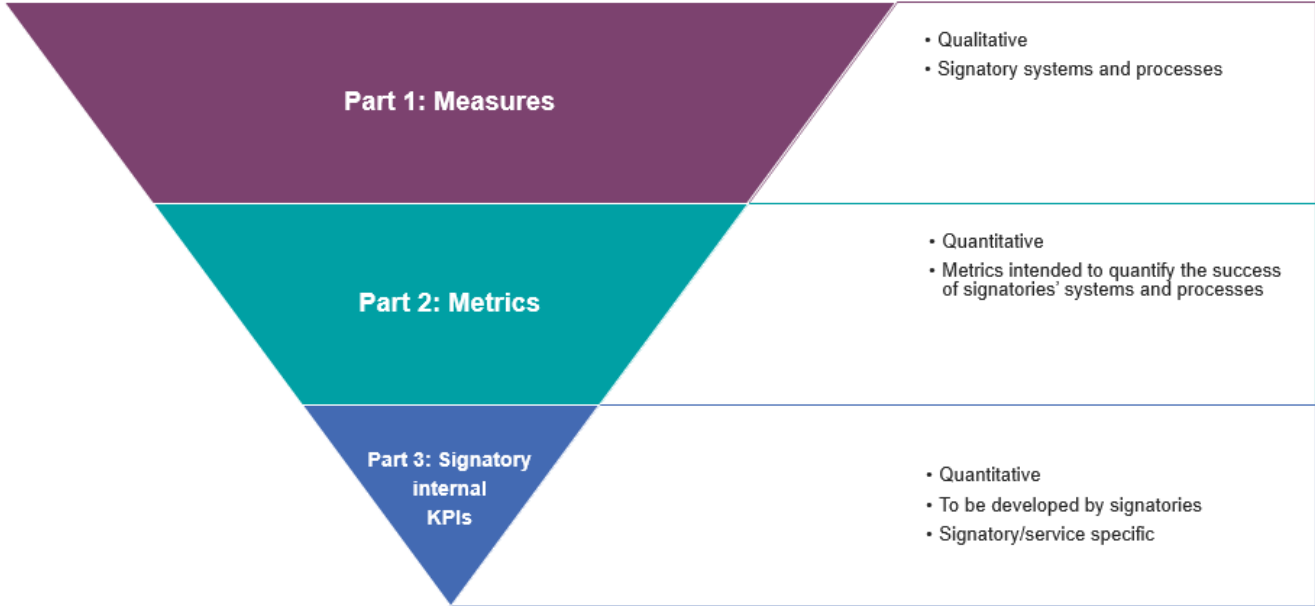
Part 1 is intended to capture qualitative information about the measures (that is, policies, systems and processes) signatories have implemented to meet Objectives 1 and 3 of the code. Part 1 is important because it captures qualitative information signatories already provide under their transparency reports but in a more consistent and comparable way. It is intended that the reporting of this information will:

- allow comparison of the various policy-based levers that signatories use to combat misinformation on their services
- provide more transparency about material changes to those policies
- create a baseline set of measures that can then be assessed via metrics.

We envisage that after their first reports against the framework, signatories will only need to report changes to these measures. Part 1 could be included as an appendix to signatories' annual transparency reports.

Part 2 of the framework establishes a set of common metrics, which are intended to quantify the effectiveness of signatories' measures. It is intended that signatories will report data against these metrics consistently each year allowing for greater comparison across years and between signatories. Information requested under Part 2 may seek data around the types and effectiveness of actions to address misinformation, the reach and impact of misinformation content, and the volume of complaints about harmful content and accounts.

It is recognised that due to the nature of diverse types of services captured by the code, not all metrics will be relevant to all services.

Part 3 of the framework is out of scope of the draft measurement framework. However, it is intended that the broader industry-wide metrics in Part 2 will allow signatories to develop internal KPIs that would measure their success. KPIs are essential for signatories to verify their progress toward the code outcomes and for industry to demonstrate the overall effectiveness of the code. KPIs are also important as they encourage services to set thresholds for success based on the nature of their offering. For example, a signatory may determine that removing 100% of violating misinformation content within 48 hours is an appropriate KPI. Another signatory with a different approach to misinformation may deem that taking action (for example, downranking, labelling) within 24 hours is an appropriate KPI.

Importantly, the framework is not intended to compel signatories to adopt specific measures to address misinformation, nor is it intended to collect information on specific harms. The primary objective of the framework is to create a more robust reporting framework that provides sufficient transparency and reassurance to the Australian public that signatories' measures are both effective at addressing misinformation, while also protecting freedom of expression.

Appendix C: Signatory report cards – May 2024 reports

Adobe	
Australian user base:	Unknown
Services covered:	All of Adobe's products
Code commitments:	Outcomes 1a, 3, 4, 6, 7
Australian data points:	Outcome 3

Strengths

- Adobe has provided detailed information, including explanations of specific projects, about the measures it employs to combat misinformation. These measures include its Content Authenticity Initiative (CAI), the Coalition for Content Provenance and Authenticity, and the use of Content Credentials in Gen AI outputs.

Weaknesses

- Adobe has not provided Australian-specific data for Outcomes 1a, 4 and 6, with the majority of data focused on global initiatives. This limits insights into the impact of its measures in Australia.
- Although one new data point has been identified, it is not Australian-specific.
- It is unclear why Adobe has continued not to opt-in to some of the code outcomes that appear relevant to the CAI such as enabling users to be better informed about the source of political advertising if that advertising features content created with Adobe products (Outcome 5).
- While good qualitative information has been provided to describe measures, limited contextual analysis has been provided about the specific data points reported. For example, explaining the significance of 200 million assets being created with Content Credentials in terms of combating misinformation would add depth.
- Case studies or examples of how these measures have effectively mitigated specific instances of misinformation could strengthen this report's narrative.

Apple	
Australian user base:	Unknown
Services covered:	Apple News
Code commitments:	Outcomes 1a, 1c, 1e, 2, 4, 6, 7
Australian data points:	Outcome 1a, 1c, 4

Strengths

- Apple has provided case studies about the quality assurance processes which underscore its public interest journalism on Apple News. Topical case studies on harm reduction include reporting on the Voice Referendum and the Israel-Hamas war.
- Apple has provided clear explanations about why it has not opted-in to certain commitments.
- Apple has provided some Australian-specific data, such as the low number of misinformation cases involving Australian publishers.

Weaknesses

- Apple has not provided Australian-specific data points for Outcomes 1e, 2 and 6, limiting quantitative insights into the impact its measures in meeting these outcomes.
- Except for complaint-related data provided for Outcome 1c, Apple has not provided year-on-year data or associated contextual commentary. This makes it difficult to assess the impact and effectiveness of measures, particularly of Outcome 1a and Outcome 2.

Google	
Australian user base:	Across all platforms: Unknown number of users YouTube: 20.4 million monthly active users (2022) Source: DPSI Report 6
Services covered:	Google Search, Google News, Google Advertising (Google Ads and Google AdSense), YouTube
Code commitments:	All code outcomes (some are only relevant to certain services)
Australian data points:	Outcomes 1a, 1c, 2, 4, 5, 6

Strengths

- Google has included consistent Australia-specific data points, such as:
  - the removal of YouTube videos uploaded from Australia that violate Community Guidelines, and specifically its Misinformation, Spam, Misleading and Scams policies
  - enforcement actions taken on AdSense pages and domains
  - creatives actioned for violating Destination Requirement policies, Inappropriate Content policies, and Misrepresentation policies in Australia
  - verified advertisers who ran election ads in Australia, and number of ads rejected
- Google has reported 6 new Australia-specific data points on appeals for YouTube and AdSense under Outcomes 1c, and 2 new Australia-specific data points on information panels and the media literacy campaign 'Hit Pause' in Australia under Outcome 4.
- Google has included an Australia-specific case study on the NSW state election and the Voice Referendum with the inclusion of 5 data points relating to ads and ad spend.
- Google has included information on updates to specific policies to combat Gen AI misinformation, including the 'About This Image' feature on Google Search, and YouTube's disclosure policy on content created by AI tools.

Weaknesses

- Google has not provided Australian-specific data points for Outcomes 1b, 1d, 1e and 3, which means the report does not provide quantitative insights into impact of Google's commitment against these outcomes.
- Google has not provided year on year data points.
- Many data points relate generally to Google's Community Guidelines, rather than being misinformation specific.
- Google reported granular data on views against violative content in its EU transparency reports but did not include this level of detail in its AU transparency report.

Meta	
Australian user base:	Facebook: 21 million monthly active users (2022) Instagram: 10 million monthly active users (2022) Threads: Unknown Source: DPSI Report 6
Services covered:	Facebook, Instagram, Threads
Code commitments:	All code outcomes
Australian data points:	Outcomes 1a, 1d, 2, 3, 4, 6

Strengths

- Meta has opted-in to 40 commitments for the next reporting period, an increase from 38 commitments for the 2023 calendar year.
- Meta has included clarifying language for its third-party fact-checking program in response to a complaint by Reset Tech Australia.
- Meta has continued to provide Australia-specific case studies, with the focus of this report on the Voice Referendum.
- Meta has included Australia-specific data and information. This includes data about Meta's actions in Australia to:
  - o act on content that violates its Harmful Health Misinformation policies
  - o display content warnings on content based on articles written by third-party fact-checking partners
  - o decisions and recommendations made by the Oversight Board.
- Meta has provided two new Australia-specific data points relating to the removal of ads that violate Misinformation policies and ads that failed to comply with its Social Issues, Elections and Politics (SIEP) ad policies.
- Meta has included information on updates to its AI policy, including the labelling of a wider range of content and disclosure requirements for advertisers.

Weaknesses

- Meta has not identified Australian-specific data points against Outcomes 1b, 1c, 1d, 1e and 5, which means the report does not provide quantitative insights into impact of Meta's commitment against these outcomes.
- Meta should indicate where measures for misinformation on Threads differ to Facebook and Instagram and report relevant data.
- Several relevant data points included in its EU transparency reports (for example, rate of reshare non-completion among the unique attempts by users to reshare a content on Facebook to feed/groups that is treated with a fact-checking label) were not included in the AU transparency report.

Microsoft	
Australian user base:	Unknown (across all platforms) LinkedIn: 5.1 million monthly active users (2022) Source: DPSI Report 6
Services covered:	Microsoft Advertising, Bing Search, Microsoft Start, LinkedIn
Code commitments:	All code outcomes (some are only relevant to certain services)
Australian data points:	1a, 3, 4

Strengths

- Microsoft has provided consistent trended data, with contextual commentary to explain changes. For instance, it details the reasons behind the decline in misinformation, attributing it to dips in traffic for topics like COVID-19, QAnon, and the Russia-Ukraine conflict.
- Microsoft has reported the date and information about Australia-specific data points, including:
  - the number of comments blocked
  - the effectiveness of AI enabled moderation in the Australian context.
- Microsoft reports data points against specific misinformation topics. For example, it provides detailed statistics on the number of Australian Takedowns of Microsoft Start comments for topics such as COVID, QAnon, and the Russia/Ukraine conflict.

Weaknesses

- Microsoft has not provided Australian-specific data points for Outcomes 1b, 1e, 2, 3, 5 and 6, limiting quantitative insights into the impact its measures in meeting these outcomes.
- Some relevant data point included in its EU transparency report (such as the number of impressions on NewsGuard) were not included in the AU transparency report.

Redbubble	
Australian user base:	Unknown ([FY2023](#) annual report showed 7.5m customers and 0.7m selling artists globally, with 6% gross transaction value from AUS/NZ)
Services covered:	Redbubble
Code commitments:	All code outcomes except Outcome 5
Australian data points:	None

Strengths

- It was positive to see Redbubble's report include some quantitative data regarding the number of accounts that are blocked or removed using their abuse detection software (data which wasn't provided last year).
- Redbubble continues to provide examples about the types of misinformation topics they screen for (such as anti-vaccine propaganda), however specific keywords would increase transparency even further.
- Redbubble provided commentary on trends observed, such as the decline in medical misinformation post-pandemic, and potential increases related to the 2024 U.S. cycle.

Weaknesses

- Redbubble has not provided Australian-specific data for Outcomes 1a, 1b, 1c, 1d, 1e, 2, 3, 4 and 6, limiting quantitative insights into the impact of its measures in meeting these outcomes.

TikTok	
Australian user base:	10 million monthly active users (2022) Source: DPSI Report 6
Services covered:	TikTok
Code commitments:	All code outcomes
Australian data points:	1a, 1d, 4, 6

Strengths

- TikTok introduced new measures to combat the spread of misinformation during the NSW state election and the Voice Referendum, with 5 specific data points reported for the latter.
- TikTok has provided some consistent Australia-specific data points, including:
    - videos removed for violating harmful misinformation policies
    - videos removed for violating Integrity and Authenticity (I&A) policies and Community Guidelines
    - percentage of I&A violations for all Community Guideline violations
- TikTok has provided one new Australia-specific data point on the Community Partner Channel under Outcome 5.
- TikTok has provided information on updates to its AI policies, including new labels for users to disclose Gen AI content.

Weaknesses

- TikTok has not provided Australian-specific data points for Outcomes 1b, 1c, 1d, 1e, 2, 3 and 5 limited quantitative insights into the impact its measures in meeting these outcomes.
- Some relevant data points (such as data on the impact of measures to reduce inauthentic behaviour such as fake likes and fake followers) included in its EU transparency report were not included in the AU transparency report.

Twitch

Australian user base:	Unknown
Services covered:	Twitch
Code commitments:	All code outcomes except Outcome 5
Australian data points:	Outcome 3

Strengths

- As Twitch's inaugural transparency report, it was positive to see the inclusion of some data and useful commentary to show how it is meeting its commitments under the code.
- Twitch has provided one Australia-specific data point on account enforcements for spam, scams and fraud in Australia.
- Twitch has provided details on how it invests in media literacy campaigns with MediaWise to develop educational materials for its active end-users.

Weaknesses

- Twitch has not provided Australian-specific data points for Outcomes 1a, 1b, 1c, 1d, 1e, 2, 4 and 6.
- The information Twitch provided about its policies to address misinformation was largely general in nature and would benefit from additional commentary to contextualise the data provided.