

A report to government on the adequacy of digital platforms' disinformation and news quality measures

JUNE 2021

Canberra

Red Building
Benjamin Offices
Chan Street
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

Gure 15

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

Copyright notice

<https://creativecommons.org/licenses/by/4.0/>

With the exception of coats of arms, logos, emblems, images, other third-party material or devices protected by a trademark, this content is made available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

We request attribution as © Commonwealth of Australia (Australian Communications and Media Authority) 2021.

All other rights are reserved.

The Australian Communications and Media Authority has undertaken reasonable enquiries to identify material owned by third parties and secure permission for its reproduction. Permission may need to be obtained from third parties to re-use their material.

Written enquiries may be sent to:

Manager, Editorial Services
PO Box 13112
Law Courts
Melbourne VIC 8010
Email: info@acma.gov.au

Contents

Executive summary	1
1. Introduction	6
2. Environmental assessment	8
3. Code development	39
4. Assessment of the code	48
5. Assessment of platform performance	64
6. Considerations for future reform	76
Appendix A: Full list of recommendations and findings	85
Appendix B: Signatory assessment reports	90
Appendix C: Timeline of key events	98
Appendix D: International regulatory approaches	123
Appendix E: Other Australian Government initiatives	134
Appendix F: Development of key performance indicators	140

Executive summary

In December 2019, as part of its response to the Australian Consumer and Competition Commission's Digital Platforms Inquiry, the Australian Government requested that digital platforms in Australia develop a voluntary code of practice to address online disinformation and news quality.

The Australian Code of Practice on Disinformation and Misinformation¹ (the code) was launched by industry association Digital Industry Group Inc (DIGI) on 22 February 2021. The code has since been adopted by 8 digital platforms – Google, Facebook, Microsoft, Twitter, TikTok, Redbubble, Apple and Adobe.

The ACMA was tasked with overseeing the development of the code and reporting to the government on the adequacy of platform measures and the broader impacts of disinformation in Australia. Our report provides new consumer research on users' experience of disinformation and misinformation on digital platforms and our assessment of the industry's code. It also provides a range of findings and a number of recommendations for consideration by the government.

The online propagation of disinformation and misinformation presents an increasing threat to Australians

Over the previous 18 months, we have seen increasing concern within the community over the 'infodemic' of online disinformation and misinformation, particularly in relation to the real-world impacts of COVID-19. The propagation of these falsehoods and conspiracies undermines public health efforts, causes harm to individuals, businesses and democratic institutions, and in some cases, incites individuals to carry out acts of violence.

To understand the scale and impacts of this issue in Australia, we undertook a mixed-methods study focused on COVID-19 misinformation. Key insights include:

- > Most adult Australians (82%) report having experienced misinformation about COVID-19 over the past 18 months. Of these, 22% of Australians report experiencing 'a lot' or 'a great deal' of misinformation online.
- > Belief in COVID-19 falsehoods or unproven claims appears to be related to high exposure to online misinformation and a lack of trust in news outlets or authoritative sources. Younger Australians are most at risk from misinformation, however there is also evidence of susceptibility among other vulnerable groups in Australian society.
- > Australians are most likely to see misinformation on larger digital platforms, like Facebook and Twitter. However, smaller private messaging apps and alternative social media services are also increasingly used to spread misinformation or conspiracies due to their less restrictive content moderation policies.
- > Misinformation typically spreads via highly emotive and engaging posts within small online conspiracy groups. These narratives are then amplified by international influencers, local public figures, and by coverage in the media. There is also some evidence of inorganic engagement and amplification, suggesting the presence of disinformation campaigns targeting Australians.
- > Many Australians are aware of platform measures to remove or label offending content but remain sceptical of platform motives and moderation decisions. There is widespread belief that addressing misinformation requires all parties –

¹ DIGI, [Australian Code of Practice on Disinformation and Misinformation](#), February 2021.

individuals, platforms and governments – to take greater responsibility to improve the online information environment and reduce potential harms.

Digital platforms have introduced a range of measures in response to the growth of disinformation and misinformation on their services

In response largely to global concerns, digital platforms have introduced measures typically based on company-wide policies including:

- > supporting third-party fact-checking organisations
- > proactively updating their policies to specifically address unique events, such as the COVID-19 pandemic and the 2020 US presidential election
- > investing in means to signal credible, relevant and authentic information
- > providing financial assistance and grants to news outlets, government and not-for-profit organisations to bolster the spread of credible information and news
- > increased detection, monitoring and enforcement action against groups and networks who use their services to spread disinformation and misinformation.

Despite platforms' mostly global approach to updating policies and implementing other actions, many measures have had an impact on Australian users.

- > In 2020, Facebook removed more than 110,000 pieces of COVID-related misinformation generated by Australian accounts.
- > Between July and December 2020, Twitter removed 50 pieces of content authored by Australian accounts for contravening its COVID-19 misleading information policy.
- > In 2020, Google blocked 101 million advertisements globally for contravening its misrepresentation policies.
- > TikTok's COVID-19 Information Hub was visited by over 292,000 Australians between November 2020 and March 2021.

The above data shows that platforms are taking proactive steps to tackle disinformation and misinformation on their products and services. The introduction of an Australian industry code builds on these actions to codify actions, improve transparency, enhance consumer protections, and implement mechanisms to monitor their effectiveness. It also provides a framework to promote stakeholder collaboration and incentivise further actions by platforms to respond to a rapidly evolving online environment.

Digital platforms have come together to develop a single outcomes-based code of practice with several important features

It is extremely positive to see industry, steered by DIGI, come together to develop a single code of practice. A single code should promote a consistent approach by platforms and provide confidence in industry to manage the range of harms associated with disinformation and misinformation.

DIGI ran a meaningful public consultation process in developing its draft code, which attracted a variety of submissions that clearly influenced subsequent changes. In particular, the scope of the code was expanded to cover misinformation as well as disinformation, a key piece of stakeholder feedback during the consultation process. The ACMA considers this is an improvement on the EU Code of Practice on Disinformation.

The code adopts an outcomes-based regulatory approach that allows a range of platforms with different services and business models to sign up to the single code. Signatories are required to sign up to the objective of 'providing safeguards against

harms that may arise from disinformation and misinformation' and may opt-in to other code objectives, such as disrupting advertising incentives and supporting strategic research.

The code also provides signatories flexibility to implement measures to counter disinformation and misinformation in proportion to the risk of potential harm. Signatories must also report annually on the range of measures they will implement to achieve the objectives and outcomes.

Importantly, the code also stresses the need to balance interventions with the need to protect users' freedom of expression, privacy, and other rights.

Our assessment identifies further improvements that should be made to the code's scope and the clarity of commitments

The ACMA has assessed the code to consider whether it has met the expectations set out by the government and has identified a range of improvements.

In our view, the scope of the code is limited by its definitions. In particular, a threshold of both 'serious' and 'imminent' harm must be reached before action is required under the code. The effect of this is that signatories could comply with the code without having to take any action on the type of information which can, over time, contribute to a range of chronic harms, such as reductions in community cohesion and a lessening of trust in public institutions.

The code should also be strengthened through an opt-out rather than opt-in model. Signatories should only be permitted to opt out of outcomes where that outcome is not relevant to their service and be required to provide justification for the decision.

The code is also limited in the types of services and products it covers. Private messaging is excluded, despite increasing concern about the propagation of disinformation and misinformation through these services, particularly when used to broadcast to large groups. Including messaging services within the code, with appropriate caveats to protect user privacy (including the content of private messages), would provide important consumer protections.

We also consider improvements to the code should be made in relation to:

- > its application to news aggregation services
- > the treatment of professional news content and paid and sponsored content
- > the weight given to news quality as a key aspect of the government's request to industry.

The ACMA is also concerned that the code does not place an obligation on individual signatories to have robust internal complaints processes. This was an area of particular concern identified in the Digital Platforms Inquiry.

The code includes commitments to establish administrative functions within 6 months of code commencement. As code administrator, DIGI will establish a compliance sub-committee, a detailed reporting guideline and a facility to address signatory non-compliance. However, these functions remain under development at the time of finalising this report. As a result, the ACMA has not been able to assess their effectiveness.

DIGI and code signatories should consider changes to the code to address the matters identified by the ACMA in its review in February 2022.

A clear and transparent measurement framework is critical to the effectiveness of a voluntary, outcomes-based regulatory model

Signatories were required to nominate their code commitments and deliver an initial report under the code, providing information and data on the measures they have adopted under the code.

Signatories' reports provide a large range of information on the actions they have taken to address disinformation, misinformation and news quality, and their investments in collaborative initiatives.

However, reports are heavily focused on platform outputs and lack systematic data or key performance indicators (KPIs) that would establish a baseline and enable the tracking of platform and industry performance against code outcomes over time. Reports also show inconsistencies in the interpretations of key code terms and in reporting formats.

Platforms should move quickly to identify KPIs specific to their services and work together to establish industry-wide KPIs to demonstrate the effectiveness of the code as an industry-wide initiative.

The ACMA recommends a number of actions by government to bolster industry self-regulatory arrangements

The ACMA considers that it is still too early to draw concrete conclusions on the overall impact or effectiveness of the code. The code administration framework – including a detailed reporting guideline and mechanism to handle complaints – is not due for completion until late August 2021. The design and implementation of these elements will be key to the overall effectiveness of the code.

Given these circumstances, continued monitoring is required and the ACMA recommends it provide government with another report on the code by the end of the 2022–23 financial year. This will provide sufficient time to assess the operation of the code administration framework and assess the impact of any changes arising from the February 2022 review of the code. As part of this report, the ACMA recommends it continues to undertake focused research on these issues.

Initial signatory reports identify challenges in obtaining relevant data on platform actions in Australia. Providing the ACMA with formal information-gathering powers (including powers to make record-keeping rules) would incentivise greater platform transparency and improve access to Australia-specific data on the effectiveness of measures to address disinformation and misinformation. Information collected could also be used to identify systemic issues across the digital platform industry and inform future ACMA research.

More formal regulatory options could be considered, particularly for platforms that choose not to participate in the code or reject the emerging consensus on the need to address disinformation and misinformation. The ACMA recommends that government provides the ACMA with reserve regulatory powers in relation to digital platforms – such as code registration powers and the ability to set standards. This would provide the government with the option to act quickly to address potential harms if platform responses are not adequate or timely.

There are also opportunities for improved collaboration between government agencies, platforms, researchers and non-government organisations on issues relating to disinformation and misinformation. The ACMA recommends that the government should consider establishing a Misinformation and Disinformation Action Group to

provide a mechanism to support future information sharing, cooperation and collaboration.

The ACMA makes 5 recommendations to the government in its report.

Recommendation 1: The government should encourage DIGI to consider the findings in this report when reviewing the code in February 2022.

Recommendation 2: The ACMA will continue to oversee the operation of the code and should report to government on its effectiveness no later than the end of the 2022-23 financial year. The ACMA should also continue to undertake relevant research to inform government on the state of disinformation and misinformation in Australia.

Recommendation 3: To incentivise greater transparency, the ACMA should be provided with formal information-gathering powers (including powers to make record keeping rules) to oversee digital platforms, including the ability to request Australia-specific data on the effectiveness of measures to address disinformation and misinformation.

Recommendation 4: The government should provide the ACMA with 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 prove inadequate, or the voluntary industry code fails.

Recommendation 5: In addition to existing monitoring capabilities, the government should consider establishing a Misinformation and Disinformation Action Group to support collaboration and information-sharing between digital platforms, government agencies, researchers and NGOs on issues relating to disinformation and misinformation.

1. Introduction

The Australian Consumer and Competition Commission (ACCC)'s Digital Platforms Inquiry (DPI) found that Australians who use digital platforms to access news and information are at risk of exposure to disinformation and misinformation.

In its December 2019 response, the Australian Government requested major digital platforms in Australia to develop a voluntary code (or codes) of conduct for disinformation and news quality.²

The ACMA was tasked with overseeing the development of the code(s) and to report to government on the adequacy of platforms' measures and the broader impacts on disinformation, with the first report due no later than June 2021. The government noted that, should the actions and responses of the platforms be found not to sufficiently respond to the concerns identified by the ACCC, it would consider the need for any further reform.

In June 2020, the ACMA produced a position paper to guide the digital industry with its code development.³ The Digital Industry Group Inc (DIGI) released a draft code in October 2020 for a 5-week public consultation. In tandem with the draft code, DIGI also hosted a roundtable discussion with targeted stakeholders to discuss the code contents and next steps.

A final code was released in February 2021, along with all 17 stakeholder submissions and a summary report.⁴ Initial code signatories were Google, Facebook, Twitter, Microsoft, Redbubble and TikTok. In May 2021, DIGI announced that Apple and Adobe had also signed up to the code and published the initial transparency reports from all 8 signatories on its website.

Alongside the code development process, agencies across the Australian Government have been working with platforms, impacted stakeholders and international counterparts on a range of related initiatives to address disinformation and misinformation (Appendix E).

Objectives of report

This report provides an examination of:

- > the broader impacts of disinformation and misinformation on digital platforms – with a specific focus on impacts in Australia (Chapter 2)
- > the process to develop the Australian Code of Practice on Disinformation and Misinformation (Chapter 3)
- > the framework and content of the code (Chapter 4)
- > platforms' measures both under the code and in addition to the code (Chapter 5)
- > considerations for further reform (Chapter 6).

² Australian Government, [Government Response and Implementation Roadmap for the Digital Platforms Inquiry](#), December 2019, pp. 6–7.

³ ACMA, [Misinformation and news quality on digital platforms in Australia: A position paper to guide code development](#), June 2020.

⁴ DIGI, [Australian Code of Practice on Disinformation and Misinformation](#), February 2021.

Methodology

The ACMA's assessment has been informed by:

- > engagement with DIGI and major platforms during the development and operation of the code
- > discussions with industry, academics, civil society, government agencies and international regulators
- > analysis of platform transparency reports
- > targeted consumer research and content analysis
- > ongoing monitoring and desktop research.

In this report, as guided by the code, we use the expression 'disinformation and misinformation' in the broadest sense to refer to the general problem of the dissemination of false, misleading and deceptive information online. Where a narrower or different meaning is intended, for example, in the discussion of academic or consumer research, this is made clear through qualifying remarks.



Defining the issue

Online disinformation and misinformation are relatively novel and dynamic phenomena and there is no established consensus on the definition of either term.

At a high level, many distinguish disinformation from misinformation on the basis of intention or behaviour. On this view, disinformation is characterised as false or misleading information created or spread to cause harm or to deceive, and misinformation as false or misleading information spread without an intention to harm or deceive. Some digital platforms treat these phenomena differently; others make no distinction.

From a regulatory perspective, the ACMA recommended in its position paper that industry take a broad view of the issue and implement a code that addresses all kinds of false, misleading or deceptive online information with the potential to cause harm. The ACMA suggested that the term misinformation more accurately reflects the full scope of the issue and the potential for harm to Australians and broader society. In the code, DIGI has chosen to maintain a distinction between disinformation and misinformation. Ultimately, however, it is not the terms used but the scope addressed by the code that is critical.

The ACMA thanks representatives of the platforms, DIGI, broader industry, academics, government agencies, impacted stakeholders and fellow international regulators for their contribution to this assessment.

2. Environmental assessment

Over the last 18 months, disinformation and misinformation has become an increasingly overt threat to Australia, and of growing concern to nearly all Australians.⁵

The COVID-19 pandemic has proven to be a lightning rod for misinformation, with governments and health officials both here and around the world recognising the urgent and ongoing need to address the ‘infodemic’ and mitigate against its real-world harms.

Through an examination of false, misleading or unproven narratives arising from the COVID-19 pandemic, this chapter provides insights into the range, spread and impact of disinformation and misinformation in Australia, providing a baseline to inform future thinking and developments across government and industry. These findings draw on desktop research, commissioned quantitative and qualitative research from the University of Canberra’s News and Media Research Centre (N&MRC),⁶ and a commissioned network analysis project from creative agency and social media consultancy We Are Social.⁷

2.1. Exposure and susceptibility

In April 2020, researchers from the N&MRC asked a representative sample of Australians about their consumption of, and engagement with, COVID-19 news, information and misinformation (wave 1). The ACMA commissioned the N&MRC to repeat and expand upon this research in late 2020/early 2021, undertaking a second survey (wave 2) as well as a series of focus groups.

Most Australians experienced misinformation about COVID-19 in 2020. When asked about the frequency of seeing news or information about the pandemic that they know or suspect to be false or misleading, 60% of Australians reported having seen some (low experience), and an additional 22% reported seeing ‘a lot’ or ‘a great deal’ of misinformation (high experience). Only 7% of respondents reported no experience of COVID-19 misinformation at all.⁸

Experiences of misinformation

Most Australians have some levels of experience of misinformation. However, those who are ‘heavy’ users of digital platforms (use of 6 or more platforms in a week) were more likely to report experiencing high levels of misinformation (‘a lot’ or ‘a great deal’).

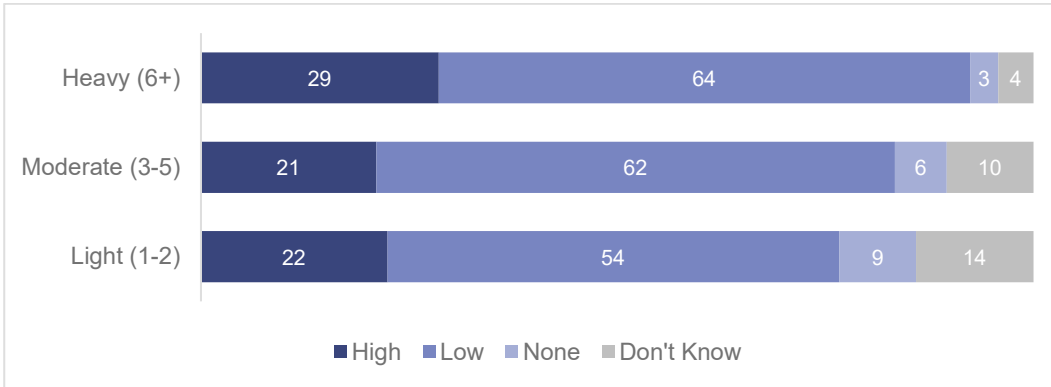
⁵ Mannheim, M. [Australia Talks data shows we don't trust Instagram influencers, but advertisers rely on them increasingly](#), ABC News, 27 May 2021.

⁶ The quantitative study consisted of a nation-wide representative survey of 2,659 adult Australians, undertaken between 19 December 2020 and 18 January 2021, based on an earlier survey undertaken by the N&MRC in April 2020. The qualitative study consisted of a series of 12 focus groups with a total of 60 participants, undertaken across February and March 2021. Participants were recruited based on a mix of demographic characteristics, geographic locations, and media habits, with a particular focus on groups that may have experiences with online misinformation but could be difficult to reach through survey research.

⁷ The network analysis project sought to examine the scale and drivers of 4 distinct online misinformation narratives (anti-vaccine, anti-5G, anti-lockdown and QAnon) in Australia over a 12-month period. This consisted of an examination of over 60,000 public conversations across Facebook, Instagram, Twitter, YouTube and Reddit, identification and analysis of 291 Australian conspiracy-driven pages and groups on Facebook and Instagram, and a manual review of misinformation narratives on TikTok and Telegram.

⁸ N&MRC, *COVID-19: Australian News & Misinformation Longitudinal Study*, 2021 [unpublished].

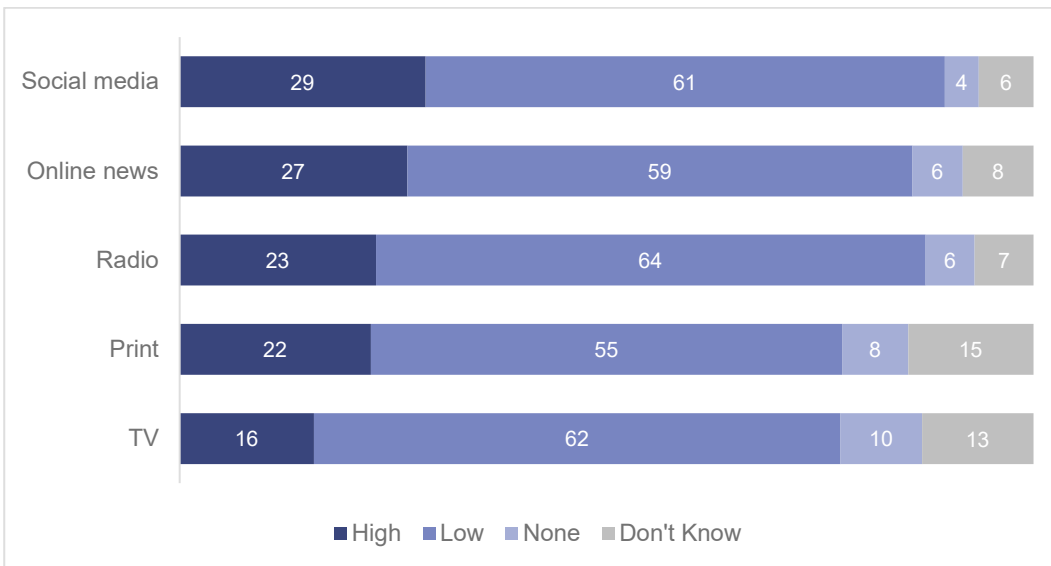
Figure 1: Experience of COVID-19 misinformation, by digital platform usage (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: High experience refers to respondents who have seen 'a lot' or 'a great deal' of misinformation. Low experience refers to respondents who have seen 'some' or 'not so much'.

Similarly, those who rely on social media as their main source of news also reported higher levels of exposure to COVID-19 misinformation than the general population (Figure 2). This was almost double the rate of those who rely on TV as their main source of news (29% and 16%, respectively).

Figure 2: Experience of COVID-19 misinformation, by main source of news (%)



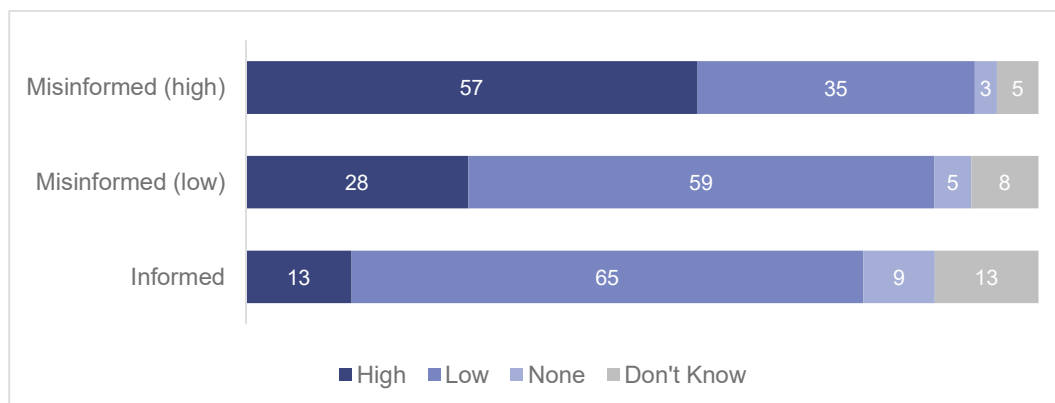
Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: High experience refers to respondents who have seen 'a lot' or 'a great deal' of misinformation. Low experience refers to respondents who have seen 'some' or 'not so much'.

This research relies on respondents both knowing and accurately self-reporting on their level of exposure to misinformation. To help address this limitation, the N&MRC also asked surveyed Australians to respond to 5 claims about COVID-19 guidelines, prevention strategies and treatments (for example, 'wearing a mask does not significantly reduce your risk of infection or spreading the virus'). Those who agreed with official advice at the time for all 5 statements were considered 'informed' (59%), while those who disagreed with 1 to 2 statements were considered 'misinformed (low)'

(30%), and those who disagreed with 3 to 5 statements were considered 'misinformed (high)' (11%).⁹

Those in the 'misinformed (high)' category were more than 4 times as likely as the 'informed' respondents to report seeing a high level of misinformation (Figure 3). As this group both report seeing more misinformation and are more likely to hold counter views to official advice, there appears to be an association between exposure to - and a potential belief in - COVID-19 falsehoods.

Figure 3: Experience of misinformation, by misinformed groups (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: High experience refers to respondents who have seen 'a lot' or 'a great deal' of misinformation. Low experience refers to respondents who have seen 'some' or 'not so much'.

It is difficult to know whether those who are highly misinformed are actually seeing more misinformation online, or if they are simply identifying any information that runs counter to their worldview as misinformation. In practice, it may be a little of both. The Pew Research Center conducted a comparable study in the United States in 2020 and found that those who mainly get their news from social media are less informed about major political events and more likely to have heard unproven claims about COVID-19, like there being a connection between 5G and the virus.¹⁰ These results suggest there is, at a minimum, a link between reliance on social media and exposure to COVID-19 misinformation.

Given the popularity of social media as a news source in Australia, these findings may be of some concern. The recently published *2021 Digital News Report: Australia* found that around half of the Australian adult population continue to access news on social media on a regular basis. A further 23% of Australians nominated social media as their main source of news – a figure that has been steadily rising each year.¹¹

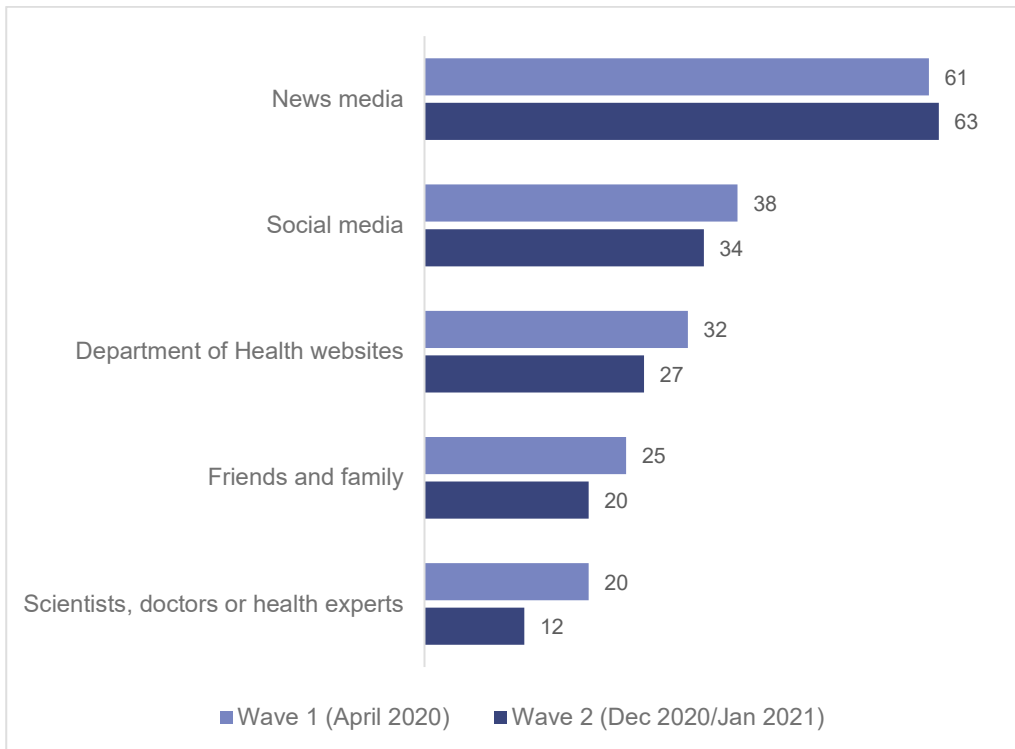
Additionally, throughout 2020, many Australians reported relying on social media as a key source of news and information about COVID-19. In both the April (wave 1) and December (wave 2) surveys, social media ranked second only to traditional news media, and was a more popular source of news and information than government websites, health experts, and friends and family.

⁹ N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

¹⁰ Mitchell, A. et al, [Americans Who Mainly Get Their News on Social Media Are Less Engaged, Less Knowledgeable](#), Pew Research Center, July 2020.

¹¹ Growing from 18% in 2019 and 21% in 2020; Park, S., Fisher, C., Lee, J., K. & McCallum, K., [Digital News Report: Australia 2021](#), News & Media Research Centre, June 2021.

Figure 4: Sources of news and information about COVID-19 (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

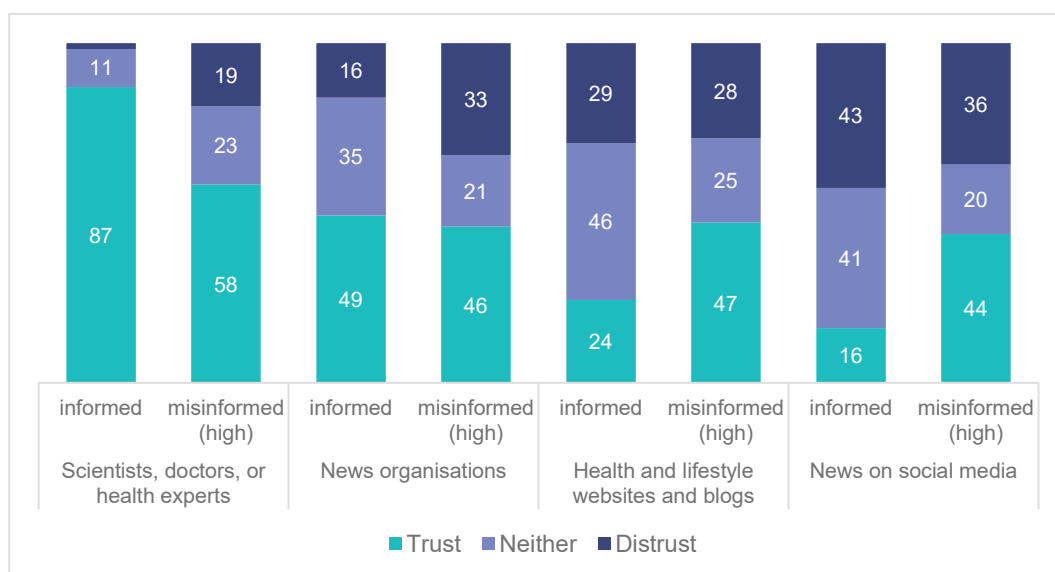
Notably, higher use or reliance on particular sources of news and information about COVID-19 does not appear to equate to higher levels of trust. Australians were most trusting of scientists, doctors or health experts (80%), and – despite its popularity – were least trusting of COVID-19 news and information found on social media (22%).¹²

A lack of trust in authoritative sources does, however, appear to be an indicator of whether or not someone believes in COVID-19 misinformation. Those who were ‘misinformed (high)’ about COVID-19 had much lower levels of trust in scientists and health professionals, and much greater trust in generally less reputable health and lifestyle websites and blogs, and news found on social media (Figure 5). This accords with other recent Australian research that shows belief in COVID-19 misinformation is closely associated with lower institutional trust.¹³

¹² N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

¹³ Pickles, K. et al, ‘COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey’, 23(1) *Journal of Medical Internet Research*, 2021.

Figure 5: Trust in news and information source, by informed and misinformed (high) groups



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: Misinformed category on this chart refers only to 'misinformed (high)' group.

Demographic differences

Those with the highest levels of experience with misinformation are more likely to be male (28%), be aged either under 25 (38%) or between 25–40 (29%) and have a high level of education (27%).¹⁴ Those who fall within this demographic profile were also much more likely to be 'very' or 'extremely' concerned about misinformation when compared to the general population.

Age appears to be one of the strongest indicators of both exposure to, and belief in, misinformation. Adult Gen Z respondents (born 1997 to 2003) were more than 3 times more likely than baby boomers (born 1946 to 1964) to rely on social media for news and information about COVID-19, as well as being 3 times more likely to fall within the 'highly misinformed' category.

In follow-up focus group discussions undertaken by the N&MRC, several parents, teachers and some Gen Z participants themselves expressed concern about younger people's overall reliance on social media, levels of exposure to misinformation, and their apparent willingness to accept conspiracy theories at face value.

[...] we're now trying to educate students who are getting educated from TikTok and openly challenge you. And it's really difficult. It's very worrying and concerning because you know that there are those kids, you'll never make a dent in them. And it wouldn't concern me if they were just ignorant. But it concerns me because they're being deliberately misinformed and are holding on to that really strongly, just like any adult who believed in QAnon.

Female, 40s, focus group #7

I think the issue is that the younger generation is a lot more impressionable and we have access to way more misinformation than probably [...] my father, my family, of that generation. They rely a lot more heavily on credible sources of information. [...] I feel like a lot of people would just rely on social media at

¹⁴ N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

our age, and that's often when, as we've mentioned before with the Facebook algorithms and the YouTube algorithms and everything like that, it just starts to propagate falsehoods, and I think we are probably the most vulnerable. Our generation.

Male, 20s, focus group #10

While the quantitative data seems to suggest that exposure and belief in misinformation is primarily an issue for younger, digitally active Australians, this is likely not a complete picture. Both the qualitative study and our desktop research suggest there are other, potentially more vulnerable, groups in society that could be exposed to equally high levels of misinformation without their knowledge.

Older Australians, for example, were more likely to report in the N&MRC survey that they did not know whether they had come across misinformation. Although this group is generally more likely to rely on traditional sources of news, like TV and print media, some focus group participants raised related concerns about their parents' changing news consumption habits, poor digital media literacy and lack of awareness of fake news when online.

I think, like, the older generations aren't educated in this, really. And so they are going down this rabbit hole [...] I'm just thinking of my parents. [Group agreement] [...] Some of the things my Dad says sometimes, I'm like 'Dad that's not right - that's not the facts, right?' [...] he also spends time on random sites online as well [...] Whereas back in the day he always tuned into SBS News or ABC News and that was the source of truth whereas all of a sudden there's all these publications and access to a lot of information.

Female, 30s, focus group #11

Some culturally and linguistically diverse (CALD) communities may also be at risk of higher exposure to online misinformation, particularly among non-English speakers. While non-English speakers were not included in our consumer research, desktop research suggests that immigrant communities rely much less heavily on mainstream media sources for news and information, and more on digital media, including foreign social media platforms and private messaging apps.¹⁵ As noted by one focus group participant, who identified as culturally Chinese, this can increase the likelihood of coming across a range of falsehoods.

So [Weibo is] kind of like Twitter, but they also can do videos and lots of news, like actual news, and some about celebrities or those kinds of gossip or whatever [...] And it is a very messy place, especially in [the] comment section. Like you can have a laugh, but in some serious matters, maybe you shouldn't believe that because all sorts of people make comments on it, so you don't know who said that, or where it comes from.

Female, 20s, focus group #8

¹⁵ See, Notley, T., Chambers, S., Park, S. and Dezuanni, M., [Adult Media Literacy in Australia: Attitudes, Experiences and Needs](#). Western Sydney University, Queensland University of Technology and University of Canberra, 2021; Zhange, S. and Chan, E., [The way misinformation travels through diaspora communities — including the Chinese diaspora — deserves more of our attention](#), First Draft website, 11 December 2020; Sun, W., [How Australia's Mandarin speakers get their news](#), *The Conversation*, 22 November 2018.

While understanding these demographic differences is important to help governments and industry design targeted interventions¹⁶, it is also important to recognise that virtually all Australians are susceptible to exposure and belief in misinformation. Based on the research outlined above, such as the clear association between trust in doctors and scientists and being 'informed' about COVID-19, improving access to authoritative sources of news and information may be an appropriate remedy to misinformation.

Finding 1: Most Australians are concerned about, and have experienced, online misinformation. Higher exposure is associated with heavy use of digital platforms, disproportionately impacting younger Australians.

Finding 2: Access to authoritative and trusted sources of news and information is an important mitigation against misinformation. Those that rely on social media as a main source of news have a greater likelihood of being misinformed about COVID-19.

2.2. Role of digital platforms

According to the N&MRC survey, over 91% of adult Australians use at least one digital platform on a weekly basis.¹⁷ The most popular platforms – Facebook, Google Search and YouTube – are each used by more than half of the population every week. These platforms are also where many Australians are consuming news and information about COVID-19 (Table 1).

Table 1: Weekly digital platform usage, by general usage, and consumption of news and information about COVID-19, Australia

#	Platform	General usage	Consumption of COVID-19 news and information
1.	Facebook	73%	46%
2.	Google Search	54%	30%
3.	YouTube	51%	22%
4.	Messenger	42%	8%
5.	Instagram	39%	16%
6.	WhatsApp	23%	6%
7.	Google News	17%	15%
8.	Twitter	16%	11%
9.	Snapchat	16%	4%
10.	LinkedIn	13%	4%
11.	Pinterest	11%	2%
12.	TikTok	9%	3%
13.	Reddit	9%	4%
14.	Apple News	6%	5%
15.	Bing	5%	3%
16.	WeChat	3%	1%

¹⁶ A recent Australian study recommended that public health messaging around COVID-19 should seek to target specific groups with higher misinformation beliefs, such as young people and those from culturally and linguistically diverse communities, and to work with these groups to ensure appropriate tonality and delivery of the message; Pickles, K. et al, '[COVID-19 Misinformation Trends in Australia: Prospective Longitudinal National Survey](#)', 23(1) *Journal of Medical Internet Research*, 2021.

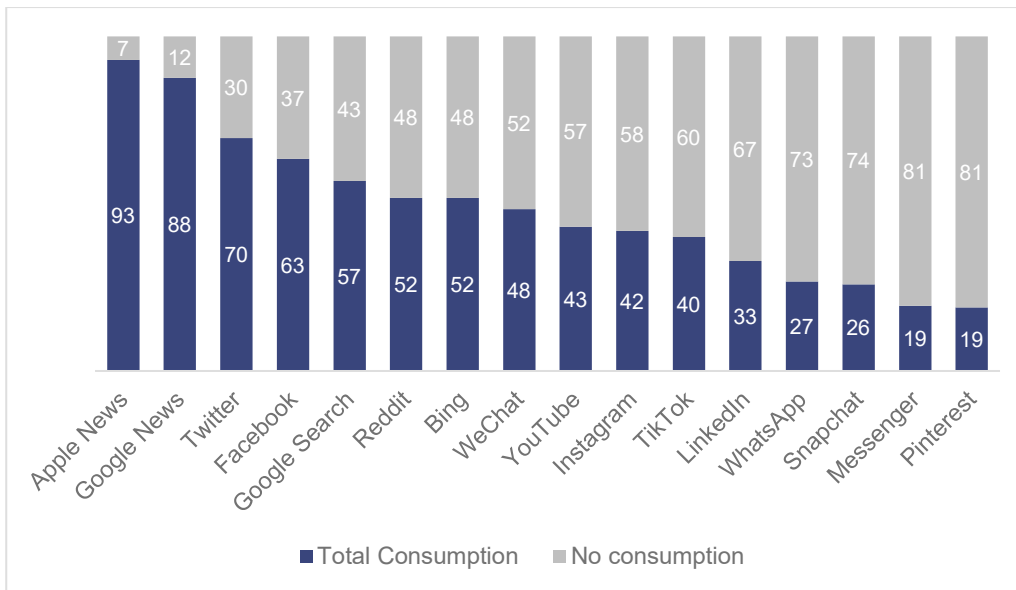
¹⁷ N&MRC, *COVID-19: Australian News & Misinformation Longitudinal Study*, 2021 [unpublished].

Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: Consumption of COVID-19 news and information includes both active and incidental consumption.

News consumption by platform users

Those who used Apple News, Google News and Twitter were the most likely to have consumed COVID-19 news and information in the last week while on the platform (Figure 6). This is consistent with the general news-oriented nature of these services, as opposed to the more social or community-oriented platforms with lower levels of reported consumption like Snapchat, Messenger and Pinterest.

Figure 6: Consumption of news and information about COVID-19, among platform users (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].
 Note: Consumption of COVID-19 news and information includes both active and incidental consumption.

Survey respondents were also asked whether their consumption was active ('I used it specifically to find news or information') or incidental ('I came across news and information while on the platform for other reasons'). The ratio of active-to-incidental consumption provides insight into what platforms users go on to find COVID-19 news and information.

Figure 7: Ratio of active-to-incidental consumption among platform users who had consumed news and information about COVID-19 (%)

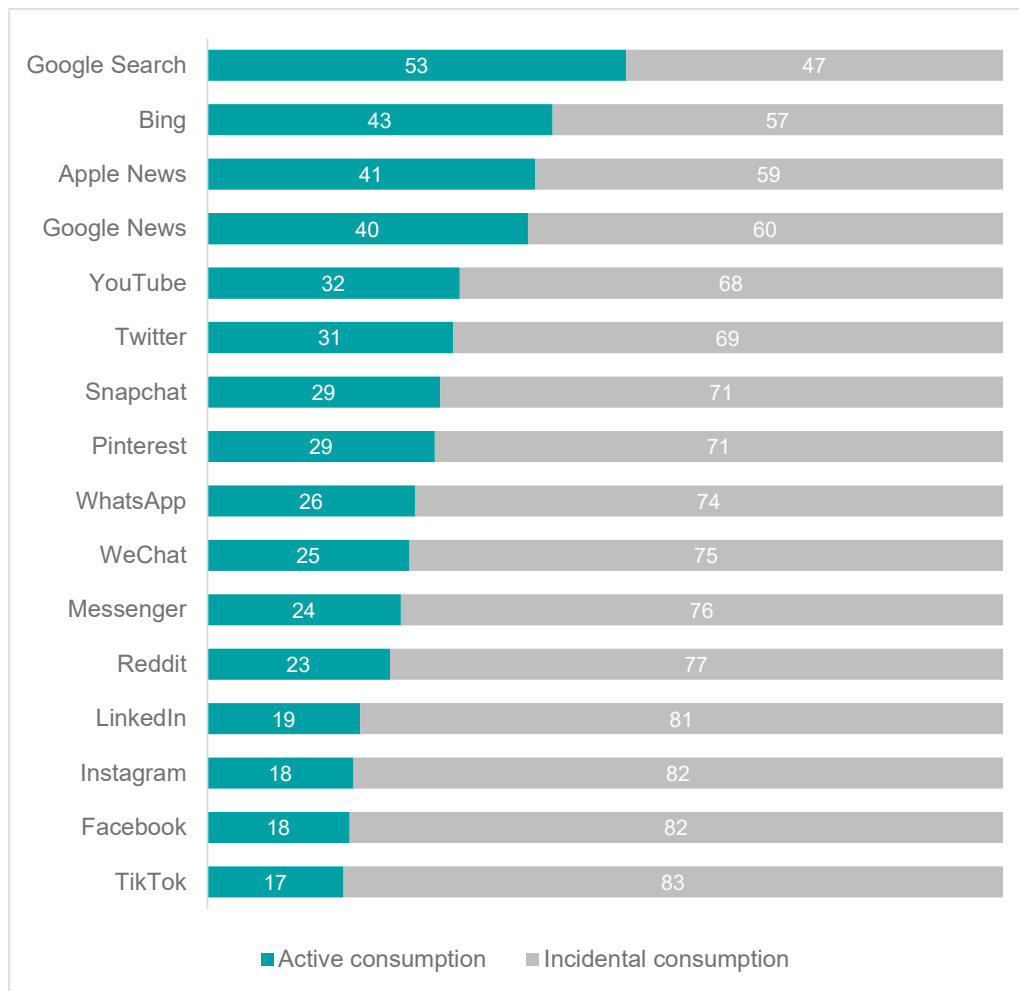


Figure 7 above shows that users of digital platforms were generally more likely to incidentally come across news and information about COVID-19 than purposefully seek it out on these services. The only exception to this was Google Search, with slightly higher levels of active to incidental consumption (53% to 47%). This result is unsurprising given that search engines respond to user-generated inquiries, noting that Google is also how most Australians access broader online resources.

If I saw something interesting [on social media], I'd be like, 'I'll have a quick read of that'. But I'm not actively seeking news, more kind of like inadvertently coming across it on social media. Unless I'm really interested about something, then I'll use Google and maybe I'd read a few different websites to get a consensus.

Male, 20s, focus group #4

At the opposite end, services like TikTok, Facebook, Instagram and LinkedIn all had very low ratios of active-to-incidental news consumption, at less than 20%. This suggests a low awareness or reliance on these platforms as a source of news and information about COVID-19.

As the most popular platform overall, and for the consumption of COVID-19 news and information (Table 1), the low rate of active-to-incidental news consumption on Facebook is particularly notable. Emerging research has found there to be a positive link between the sharing of misinformation and the attention levels of users on social media.¹⁸ As users who passively or incidentally consume news and information are less likely to focus on the accuracy of the content posted, platforms with lower ratios of active-to-incidental news consumption may be at higher risk of facilitating the propagation of misinformation.

I don't go to Facebook to look at information, I just go to socialise.
But the problem is that these media outlets exist on Facebook, and that's how I end up getting information.

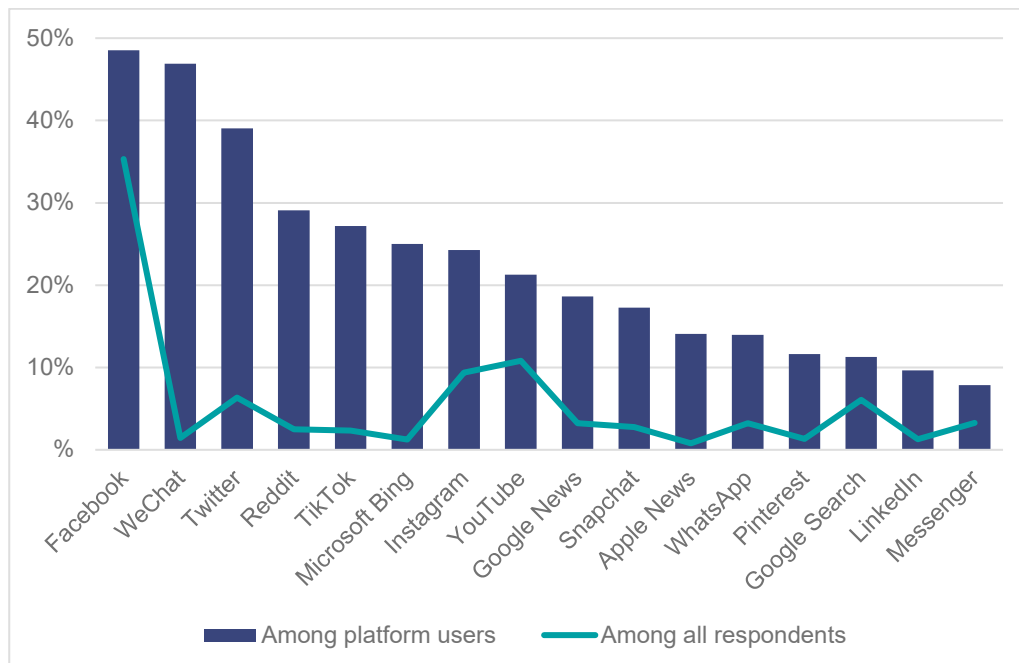
Male, 20s, focus group #8

Misinformation by platform

Given the constantly shifting nature of misinformation, difficulties in assessing falsehoods, and the challenges in accessing relevant data, it is not possible to quantify the true scale and volume of misinformation in Australia. However, consumer surveys and open-source network analysis can help provide insights into which platforms are at higher risk and help provide a baseline for future assessments.

We asked Australians who reported seeing COVID-19 misinformation to identify on which digital platforms they had seen it. Figure 8 below shows the split among all survey respondents (green line), as well as the breakdown only among the users of that platform (blue column).

Figure 8: Reported experience of COVID-19 misinformation by platform



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

¹⁸ Pennycook, G., et al., 'Shifting attention to accuracy can reduce misinformation online', *Nature*, 592, 590-595, 2021.

Facebook had the largest share of reported COVID-19 misinformation, with over a third of all Australians who had reported seeing COVID-19 misinformation (35%) identifying Facebook as the platform they saw it on.

This ranking holds true even when accounting for its much higher usage than any other platforms. Among the 73% of Facebook users in Australia, just under 50% report seeing misinformation. This suggests that Facebook has a disproportionately high impact on the volume of online misinformation present in Australia.

WeChat has the second highest level of reported misinformation among its respective users (47%). This is consistent with reporting on the widespread availability of misinformation on the platform and is a notable finding given WeChat’s popularity in certain CALD communities.¹⁹ Unlike Facebook, however, weekly use of WeChat is relatively low within the overall Australian community (3%).²⁰ As such, platforms that are more widely used like Twitter, Instagram and YouTube are likely to have a greater impact on the total volume of misinformation in Australia.

These survey results are broadly consistent with the findings from our network analysis. Our researchers, We Are Social, compiled and analysed a sample of publicly available Australian conversations posted on social media related to 4 misinformation narratives (anti-5G, anti-vax, anti-lockdown and QAnon) over a 12-month period. This was based on a list of popular keywords used by supporters of each misinformation narrative (such as #5gmindcontrol, #covid19vaccineexposed, #masksdontwork and #qaustralia), across Facebook, Twitter, Instagram, Reddit and YouTube. An overview of these findings is at Table 2.

Table 2: Volume of mentions and interactions relating to key misinformation narratives in Australia, by platform, April 2020 – March 2021

Platform	Mentions	Likes	Comments	Shares	Potential impressions
Facebook	16,527	2,134,912	711,643	782,527	430,777,798
Twitter	45,149	23,459	36,152	80,789	121,694,338
Instagram	14,270	2,139,991	183,939	N/A	114,996,577
Reddit	461	17,467	7,791	N/A	30,193,932
YouTube	216	10,396	3,602	8,236	3,130,516

Source: We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021 [unpublished]*

¹⁹ See, for example, Xiao, B., T. Aualiitia, N. Salim and S. Yang, ‘[Misinformation about COVID vaccines is putting Australia’s diverse communities at risk, experts say](#)’ ABC News, 4 March 2021.

²⁰ Noting the small number of respondents who used WeChat in the past week (n=81), we would also expect a higher margin of error in these results.

This research highlights how a relatively small number of posts can generate high levels of sharing and engagement, reaching significant numbers of people. Across the 5 major social media platforms examined, Twitter had the highest overall number of conversations that mentioned misinformation keywords (Table 2). Facebook and Instagram ranked second and third respectively, but posts on these platforms had much higher levels of overall engagement and potential impressions than those on Twitter, driven by accounts with larger audiences.

The We Are Social research does not constitute an exhaustive examination of misinformation in Australia. Rather, the intent of this exercise was to provide high-level trends and insights into selected misinformation narratives over time. The analysis is heavily dependent on the list of chosen keywords for each narrative²¹, and subject to broader limitations around the availability and use of data. For example, some platforms do not provide access to geographic information or include information on content that has previously been removed. This may be one reason for YouTube's low number of mentions, despite appearing to play a significant role in conspiracy-driven online communities.²²

Finally, social media platforms prevent researchers from examining any conversations within private or closed groups or communities, irrespective of the size or reach of these groups. While recognising the privacy rationale behind this decision, this creates a significant gap in our understanding, and is an area of concern for many social media researchers due in part to credibility amplification from false content being shared by friends and family.²³

Alternative platforms for misinformation

With the strengthening of platform moderation activities by some platforms throughout 2020 and the first half of 2021 (see Appendix C), an important trend in the propagation of misinformation has been the migration of some conspiracy-driven individuals and communities from mainstream sites like Facebook and Twitter towards smaller, free-speech oriented platforms with less rigorous content moderation policies.

In the We Are Social sample of 200 conspiracy-driven Facebook groups and pages, there were 4,479 mentions of alternative social networks, with many of these posts inviting users to join alternative social networks (Figure 9).²⁴ Among the range of alternative platforms used by conspiracy-driven communities in Australia – including Gab, Parler, Rumble and MeWe – the most popular platform by total number of references is Telegram.

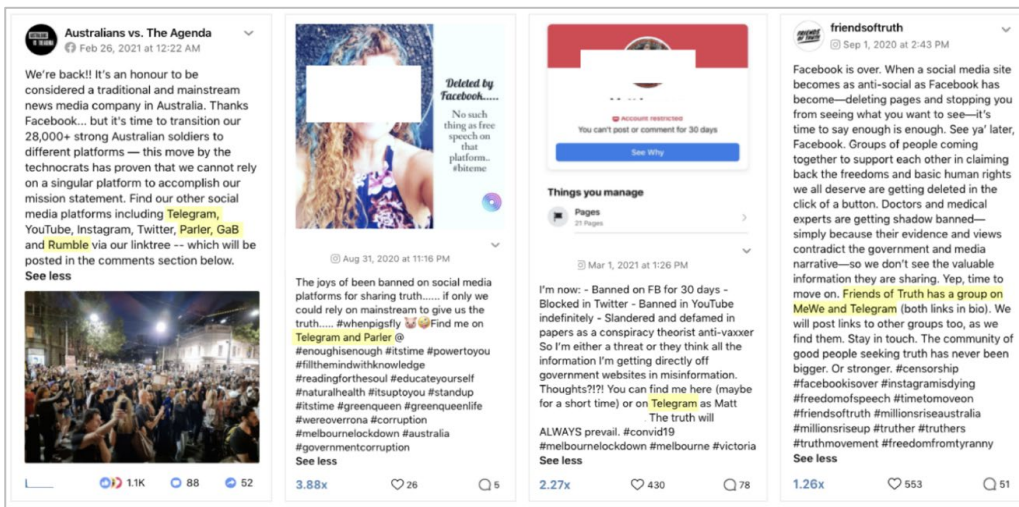
²¹ This approach may also capture a number of conversations that use identified misinformation keywords but do not appear to be spreading misinformation or form part of a broader misinformation narrative – including commentary, criticism and satire. We Are Social estimates this could represent up to 30% of the sample; We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

²² For example, YouTube represented 16% of the top 50 link sources among the sample of 200 conspiracy-driven Facebook group and page accounts; We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

²³ See, for example, Tantuco, V., [On Facebook's messaging apps, false information spreads undetected, unchecked](#), *EU Disinfo Lab*, 12 February 2021.

²⁴ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

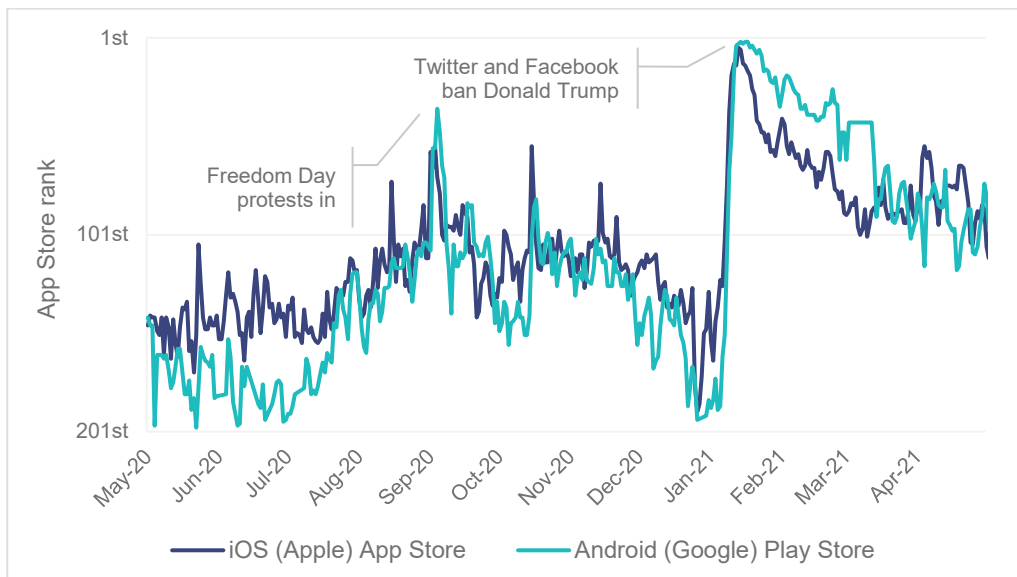
Figure 9: Examples of Facebook posts from conspiracy-driven communities, inviting followers to join alternative platforms



Source: *We Are Social, Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021 [unpublished].*

In Australia, use of Telegram has grown rapidly. App store rankings for Telegram first increased noticeably in early September 2020, likely related to the organisation of the anti-lockdown ‘Freedom Day’ rallies. Australian downloads of the Telegram app then peaked in early January 2021 for both Apple iPhone (reaching #6 in the iOS App Store) and Android devices (reaching #3 in the Google Play Store). This was likely tied to a number of concurrent events, including the US Capitol riots, the removal of Parler from both app stores, and Donald Trump’s ban from Twitter and Facebook (Figure 10).

Figure 10: Telegram app store rankings in Australia, iOS and Android



Source: *SimilarWeb, App store rankings, 1 May 2020 to 30 April 2021.*

Conspiracy-driven Telegram channels gain subscribers through the circulation of channel lists, allowing for discoverability of channels from similar groups and like-minded individuals. While Telegram conversations were not included in the We Are Social quantitative analysis, these lists allowed researchers to identify a separate sample of the most popular conspiracy-driven Telegram channels in Australia.

Former television chef Pete Evans appears to have the most popular conspiracy-driven Telegram channel in Australia, posting an average of 12 posts per day to his 35,000 followers, with many of his posts containing unproven or objectively false claims about COVID-19 cures and QAnon conspiracies.

While this pales in comparison to the number of followers Pete Evans had on Facebook and Instagram before he was banned from these platforms in February 2021, his Telegram channel has more subscribers than almost all of the local conspiracy-driven Facebook communities and pages identified by We Are Social.²⁵ As discussed further below, this highlights the role of local influencers in amplifying misinformation narratives.

Finding 3: Given its nature and the ongoing challenges in accessing relevant data, the true scale and volume of misinformation in Australia is currently unknown.

Finding 4: Australians report seeing the most amount of misinformation on large platforms such as Facebook and Twitter. However, private messaging services and smaller platforms with less strict content moderation policies, like Telegram, are also being embraced by conspiracy-oriented communities.

2.3. Sources and amplification

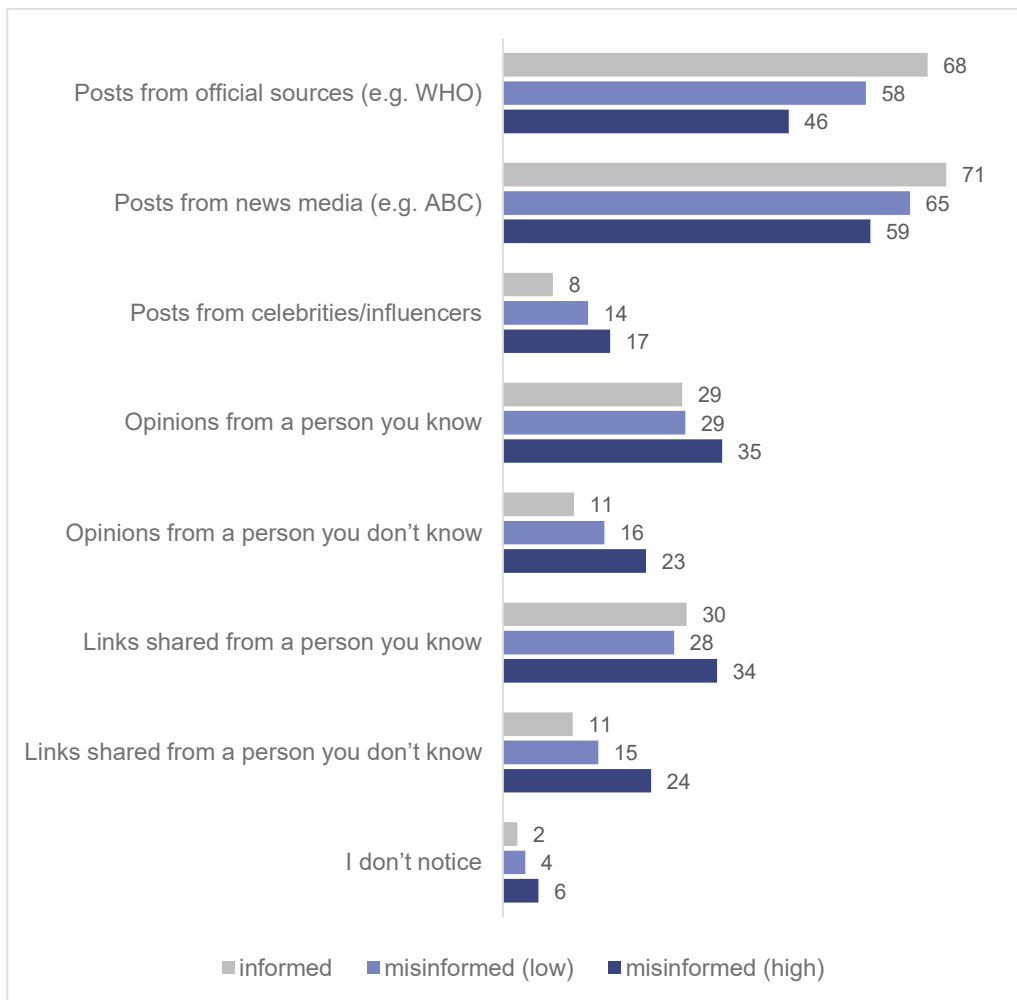
To assess overall attentiveness and understand more about the origins of online misinformation, we asked Australians who had seen COVID-19 news and information on social media whether they could recall or identify the source of the news or information.

Social media posts from 'official sources' (such as the Australian Government or the World Health Organisation) and 'news media' (such as the ABC or *The Australian*) were the most popular sources of COVID-19 news and information overall. However, those who were 'highly misinformed' about COVID-19 were more likely than the 'informed' respondents to get their information from less reputable sources, like celebrities and social media influencers, links posted by people they know, or links posted by strangers.²⁶

²⁵ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia*, May 2021 [unpublished].

²⁶ The Digital News Report: Australia 2021 also found important differences in news sources between platforms, with social media personalities and influencers (21%) the second most common news source among Instagram users, behind mainstream news outlets or journalists (23%); Park, S., Fisher, C., Lee, J., K. & McCallum, K., [Digital News Report: Australia 2021](#), News & Media Research Centre, June 2021.

Figure 11: Sources of news and information about COVID-19 while on social media (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

Influencers and conspiratorial convergence

Conspiracy theories tend to start small and remain mostly insular, shared within communities of people that already agree with one another, forming echo-chambers that reinforce a particular worldview.²⁷ However, conspiratorial content can quickly propagate to a wider audience via super-spreaders, who can draw audiences from distinct communities into shared misinformation narratives.²⁸

In a recent large-scale internal study of vaccine-hesitant content on its platform, Facebook's data scientists identified half of this content originated from just 10 out of 638 population segments. Among the population segment with the most vaccine hesitancy, only 111 users were responsible for half of all vaccine-hesitant content.²⁹

²⁷ Sunstein, C. and Vermule, A., '[Symposium on Conspiracy Theories: Causes and Cures](#)', *Journal of Political Philosophy*, 17(2), 202-227, November 2009.

²⁸ Starbird, K. et al, '[Ecosystem or Echo-System? Exploring Content Sharing across Alternative Media Domains](#)', *Proceedings of the International AAAI Conference on Web and Social Media*, 2018.

²⁹ Dwoskin, E, '[Massive Facebook study on users' doubt in vaccines finds a small group appears to play a big role in pushing the scepticism](#)', *The Washington Post*, 15 March 2021.

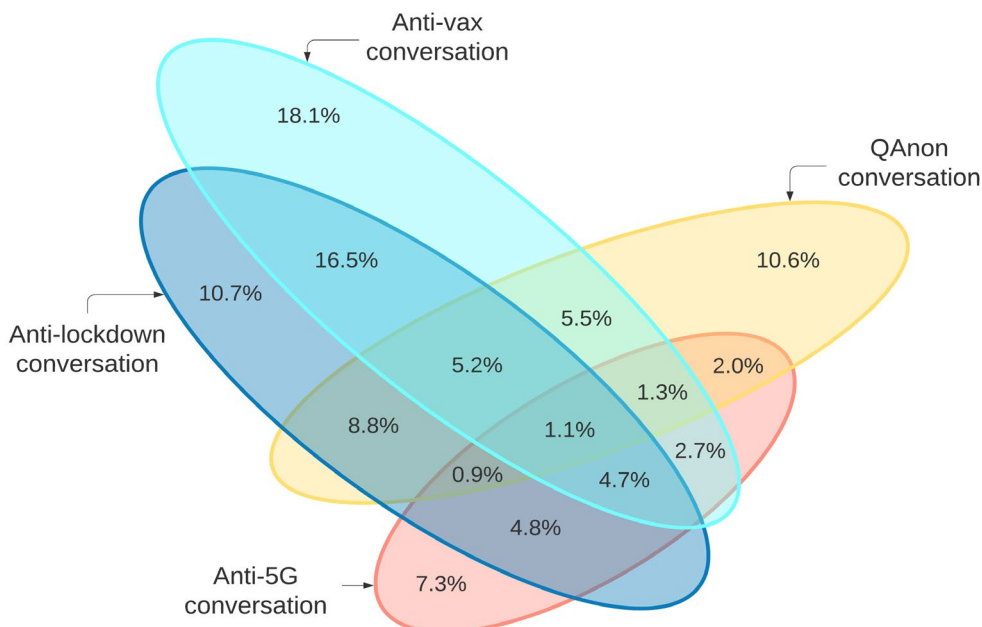
Throughout the pandemic, we have similarly seen a relatively small number of Australian celebrities, sporting figures, politicians and prominent online influencers exerting an outsized influence over COVID-19 misinformation narratives, while also growing loyal, vocal and highly engaged groups of online supporters.

In our commissioned research, We Are Social compiled a list of the top 20 Australian influencers sharing misinformation narratives, based on their total number of interactions. Heading this list was celebrity chef Pete Evans, followed by Federal MP Craig Kelly, and prominent anti-vaccine campaigner Taylor Winterstein.³⁰

COVID-19 conspiracy theories have proven to be particularly pervasive because they have brought together people from different backgrounds and with very different concerns. This includes members of the holistic health community, who are distrustful of pharmaceutical companies and vaccines, anti-authoritarian groups who see lockdowns as an attack on the freedoms of the individual, and followers of QAnon who believe COVID-19 is a government cover-up.

There is considerable overlap between the 4 misinformation narratives examined by We Are Social. From the list of the top 20 local influencers discussed above, 12 (60%) had posted on topics relating to all 4 narratives. Similarly, more than half of the posts in the We Are Social sample (54%) include misinformation keywords from at least 2 narratives. The largest overlap was between anti-lockdown and anti-vaccine conversations (16.5%), followed by anti-lockdown and QAnon (8.8%). All 4 narratives were referenced in 1.1% of posts (Figure 12).

Figure 12: Share of conversation by selected narrative within selected conspiracy-driven groups and accounts, April 2020 to April 2021



Source: We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

Note: Based on share of conversation across a sample of 100 Facebook groups, 100 Facebook pages and 91 Instagram accounts. Diagram is illustrative and not proportionate. Does not equal 100% due to rounding.

³⁰ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

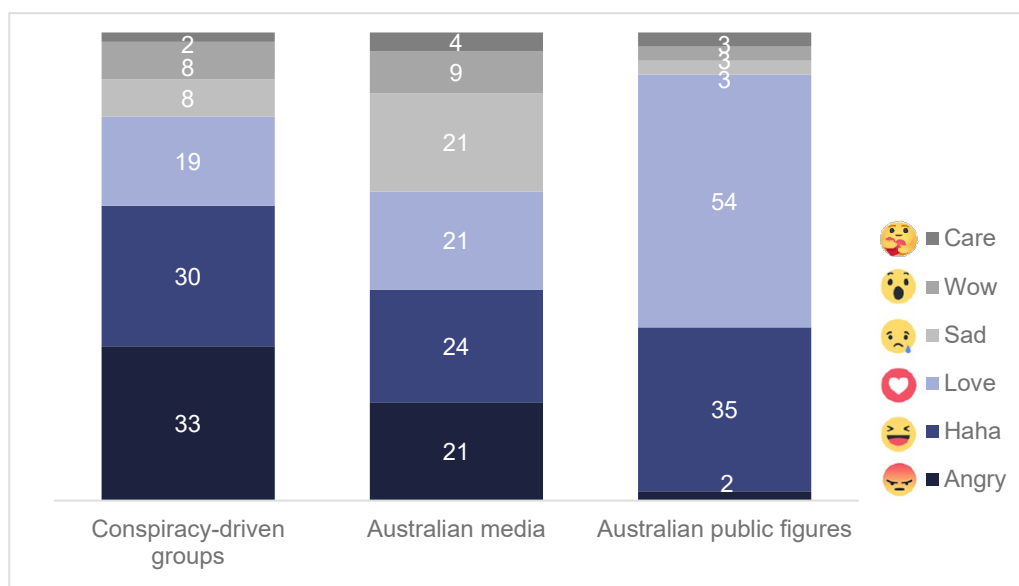
Amplification techniques

Online misinformation is often characterised as a ‘rabbit hole’. The user journey may start innocuously, such as by engaging with a piece of online content raising legitimate concerns around vaccine safety. However, surrounding discussions often contain pathways to misinformation, such as links to anti-vaccine conspiracy websites, or invitations to join online communities of like-minded people.

There are a range of social and cognitive biases that drive people to engage with misinformation, such as in-grouping, political fragmentation, and identity-based conflict. Research suggests people may be driven to engage with misinformation on ideological grounds or to defend their worldview, even if they do not believe the content.³¹

Content created for these communities is designed with virality in mind, using a variety of techniques to promote engagement. One approach is to post material that fuels outrage, such as unverifiable but highly intimate and emotive first-hand anecdotes.³² Our research shows that the most common reaction among users engaging with conspiracy-driven Facebook groups is ‘anger’ (Figure 13). By contrast, the most common reaction to Facebook posts from mainstream Australian media outlets is ‘Haha’, and the most common reaction to posts from the top 745 most popular Australian public figures is ‘Love’.

Figure 13: Share of Facebook reactions within conspiracy-driven groups, Australian media and public figures, May 2020 to April 2021 (%)



Note: ‘Conspiracy-driven groups’ consists of 100 Facebook groups from the We Are Social sample. ‘Australian media’ represents the Facebook pages of 49 mainstream media outlets in Australia. ‘Australian public figures’ represents the 745 most-followed Australians on Facebook.

Source: We Are Social, Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021 [unpublished].

Another way to promote engagement and amplification of misinformation is by building and strengthening a sense of community and shared belief. Members of these communities see themselves as ‘free thinkers’ or ‘truth seekers’, who avoid

³¹ See, for example, Nefes. T. S., ‘[The impacts of the Turkish government’s conspiratorial framing of the Gezi Park protests](#)’, *Social Movement Studies*, April 2017.

³² Glaser, A. and Zadrozny, B., ‘[Distancing from the vaccinated: Viral anti-vaccine infertility misinfo reaches new extremes](#)’, NBC News, 14 May 2021.

mainstream news and rely on their communities for alternative news.³³ The use of visual memes, symbolism and coded language in posts helps create a shared identity, built around secret or hidden knowledge that the public is not privy to, or that people in authority are trying to hide.³⁴

In this regard, some content moderation techniques adopted by digital platforms can be ineffective or counterproductive in addressing misinformation.³⁵ Content removal or de-platforming feeds into the general belief that platforms are involved in a deep-state 'cover-up'. It also encourages members of conspiracy-driven communities to take steps to pro-actively avoid detection or automated content moderation tools. We Are Social found widespread use of intentionally misspelling keywords in posts, such as 'v8ccine' and 'vackseen'.³⁶

Widespread content moderation by the platforms may also drive these conversations further underground, by encouraging mass migrations to smaller alternative social media or encrypted messaging apps. This makes it harder for governments, researchers and platforms themselves to monitor and address harmful disinformation and misinformation.

Role of news stories and the media itself

To further increase engagement and amplification, those seeking to spread misinformation often reinterpret and leverage current events, political announcements and other news of the day.

It is common for Australian conspiracy-driven Facebook communities to link to articles and videos from mainstream Australian news sources (for example, Sky News, ABC) and provide supporting commentary that is either critical of the news story or suggest that it provides evidence or support of a broader misinformation narrative. Of the top 50 links shared on these group and page accounts, 40% were to mainstream news websites, while another 23% were alternative news websites or blogs from unreliable or untrusted sources (Figure 14).

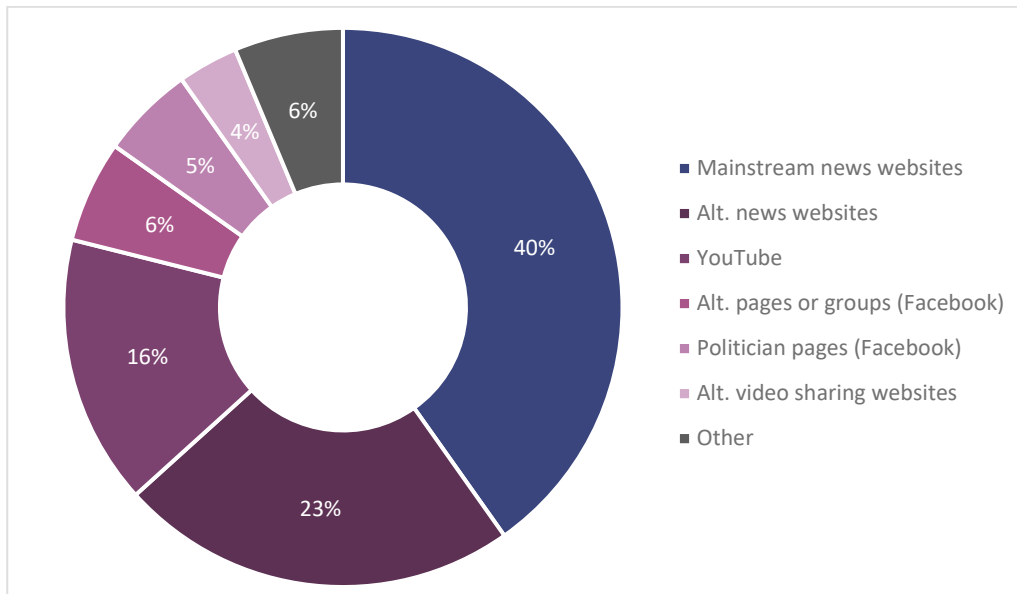
³³ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia*, May 2021 [unpublished].

³⁴ Marwick, A., '[Why do people share fake news? A sociotechnical model of media effects](#)', 2 *Georgetown Law Technology Review*, 474 (2018). Researchers also note evidence of diversity and dissent in opinion within conspiracy-driven communities, which is removed when content like memes shift across platforms, decontextualising and amplifying false narratives; Krafft, P. M. and Donovan, J., '[Disinformation by Design: The Use of Evidence Collages and Platform Filtering in a Media Manipulation Campaign](#)', *Political Communication*, 37:2, 194-214.

³⁵ Smith, R. et al, '[Under the surface: Covid-19 vaccine narratives, misinformation and data deficits on social media](#)', First Draft, November 2020, p. 14.

³⁶ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia*, May 2021 [unpublished].

Figure 14: Top 50 Facebook link sources from 200 Australian conspiracy-driven Facebook group and page accounts, by source category



Source: We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

In discussions with the ACMA, Australian news organisations have noted that they are cognisant of this issue, particularly in regard to potentially misleading article headlines and thumbnails used for video clips. Some outlets report they are now taking more active steps to reduce the likelihood that their news content is misused or taken out of context, like increasing social media training for journalists and other content creators.

Misinformation narratives can also be amplified by news media outlets themselves. They can direct considerable attention to falsehoods and help these conversations find a much wider audience, particularly where a story involves a celebrity or public figure.³⁷ Focus group participants discussed the importance of the news media in addressing or fact-checking misinformation but were also concerned about the impact of the media in amplifying misinformation. Participants from separate groups also made unprompted comments on the state of media diversity in Australia when asked about who’s responsibility it was for combatting misinformation.

When I see misinformation or just blatant conspiracy theories being perpetuated online and [then] actually make their way into the sort of mainstream media discussion, and people are being exposed to them on a wide basis, that makes me concerned. [...] I mean, they should be addressed, they definitely need to be addressed by the mainstream media and shut down, but it’s concerning that they’re reaching millions and millions of people.

Male, 20s, focus group #10

I think obviously it’s much, much easier for misinformation to be spread if there’s a limited diversity of news sources.

Female, 20s, focus group #10

³⁷ See, for example, Evanega, S. et al., [‘Coronavirus misinformation: quantifying sources and themes in the COVID-19 ‘infodemic’](#), the Cornell Alliance for Science, 2020.

I think the biggest thing is making sure there's media diversity so you get information from all sorts of perspectives, instead of just being controlled in a centralized way.

Male, 20s, Focus Group #4

International influence

Many misinformation narratives originate overseas but evolve or adapt to local audiences and domestic issues. The best example is the QAnon conspiracy.

Originating as a series of cryptic messages on the imageboard 4chan in 2017, QAnon was built on debunked far-right claims that Democratic political figures like Hillary Clinton belonged to a cabal of Satan-worshipping paedophiles. Over time, this conspiracy grew to include other 'powerful elites' like Bill Gates and Pope Francis, as well as encompass a broader range of deep-state and anti-globalist rhetoric that could be applied in local contexts. During Melbourne's 2020 lockdown, for example, Australian QAnon followers heavily promoted the theory that the state government lockdown was a cover to traffic stolen children through secret tunnels under the city.³⁸

In 2020, Australia was considered the 4th most active QAnon country in the world, responsible for approximately 2% of the global QAnon-related Twitter conversations.³⁹ QAnon was also the most popular of the 4 Australian misinformation narratives We Are Social examined over the 12-month period to March 2021, representing 31% of the total conversations in the broader sample.⁴⁰ In the N&MRC consumer research, several of the focus group participants shared QAnon-related anecdotes, with one man talking about how his daughter had become a QAnon believer.

[...] when I hear her talking to her friends who I know, it's like listening to, I don't know, Nazi propaganda. It's just like where the frig do they find this stuff out from? It's all QAnon crap – we were talking about Trump a few months ago – and she goes 'oh he's done some good things' and I said, 'what?' and she said, 'oh against the paedophiles' [group laughter] [...] working out of a pizza shop wherever it is in Washington somewhere, Hilary Clinton and eating corpses now. It's just nuts.

Male, 60s, focus group #6

International supporters of misinformation narratives contribute to, and engage with, local conspiracy-driven networks and communities. We Are Social estimate that almost 60% of the identified misinformation conversations within their sample contained global rather than local themes. Among Australian conspiracy-driven communities on Facebook and Instagram, former US President Donald Trump was mentioned 3 times more than Australian Prime Minister Scott Morrison, and 17% of the Facebook pages in the We Are Social sample had international administrators, mostly from the US, Canada and Israel.⁴¹

Disinformation campaigns

In addition to promotion from genuine supporters, some conversations within Australian misinformation communities appear to be spread inorganically, such as via

³⁸ Kolankiewicz, V. '[What lies beneath: tunnels for trafficking, or just a subterranean service? Time to rescue these spaces from the conspiracists](#)' *The Conversation*, 14 September 2020.

³⁹ Gallagher, A. et al., '[Key trends in QAnon activity since 2017](#)', ISD, 2020.

⁴⁰ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia*, May 2021 [unpublished].

⁴¹ *ibid.*

the use of bots or fake users. This suggests the presence of disinformation campaigns, typically orchestrated by bad actors seeking financial gain, or by foreign governments or other entities seeking to intentionally cause social harm by undermining trust in a democratic process, the breakdown of community cohesion, or the destabilisation of local institutions.

Facebook has publicly identified and removed 4 coordinated inauthentic behaviour (CIB) networks directly relevant to Australia – 2 where Australia was a target of foreign CIB networks, one where Australia was a country of origin in a CIB network targeting foreigners, and one instance of domestic CIB.⁴²

Table 3: Facebook disclosure of known CIB networks of relevance to Australia

Date of public disclosure	Domestic/ Foreign	Overview
8 March 2019	Domestic	Facebook suspended the personal account of a candidate running for NSW parliament after an aide set up multiple fake accounts to smear a political rival. ⁴³
26 March 2019	Both	Australia was one of several countries targeted by a CIB network, originating in Macedonia and Kosovo. This network used fake accounts purporting to represent local political communities and posted about religious and political topics. ⁴⁴
3 October 2019	Both	Australia was one of several countries targeted by a CIB network using fake accounts and localised content to artificially increase engagement and promote content related to the UAE. Facebook identified links to marketing firms operating out of the UAE and Egypt. ⁴⁵
6 August 2020	Both	Facebook removed a CIB network that operated from multiple regions, including Australia, linked to the digital media outlet Truthmedia. ⁴⁶

Source: Facebook, various CIB reports.

It is difficult to identify inauthentic coordinated activity, and even more difficult to identify the sources behind it. Nevertheless, it is almost certain that Australians have been the target of many more online disinformation campaigns than those published by Facebook – particularly over the last 18 months.

In January 2020, for example, researchers at Queensland University of Technology (QUT) discovered evidence of Twitter bots being used to amplify the narrative that Australia’s 2019/20 bushfires had been deliberately lit.⁴⁷ In July 2020, researchers from the Australian Strategic Policy Institute (ASPI) found reports of deaths from an

⁴² Facebook, [Threat Report: The State of Influence Operations 2017-2020](#), May 2021.

⁴³ For this matter, Facebook provided comments via local media reporting rather than publish details on its website; Millington, B., [‘Port Stephens Liberal candidate Jaimie Abbott linked to trolling from fake Facebook accounts’](#), ABC News, 9 March 2019.

⁴⁴ Facebook, [Removing Coordinated Inauthentic Behavior from Iran, Russia, Macedonia and Kosovo](#), 26 March 2019.

⁴⁵ Facebook, [Removing Coordinated Inauthentic Behavior in UAE, Nigeria, Indonesia and Egypt](#), 3 October 2019.

⁴⁶ Facebook, [July 2020 Coordinated Inauthentic Behavior Report](#), 6 August 2020.

⁴⁷ Graham, T., and Keller, T., [‘Bushfires, bots and arson claims: Australia flung in the global disinformation spotlight’](#), *The Conversation* 10 January 2020.

entirely fictitious US-led vaccine trial in Ukraine appearing on prominent Australian anti-vaccination Facebook groups, days after Russia announced plans to develop its own vaccine.⁴⁸ There have also been examples of disinformation used by state-aligned actors to publicly criticise Australia, such as the well-publicised tweet from a Chinese foreign ministry spokesperson in November 2020, publishing a doctored image of an Australian soldier holding a knife to the throat of a child.⁴⁹

Disinformation poses an ongoing threat to Australia. In addition to bots and troll farms, a range of new technologies, underpinned by advancements in AI and machine learning, continue to evolve and provide new tools for bad actors to intentionally spread harmful falsehoods.⁵⁰

A number of initiatives are underway across the Australian Government to monitor and respond appropriately to broader disinformation campaigns that can cause damage to Australia's interests or reputation (see Appendix E). However, it is also necessary to recognise the range of downstream harms that can occur once this type of content is widely shared by ordinary users. There is an ongoing role for government in monitoring harmful disinformation and misinformation campaigns, as well as a broader coordination role in bringing together stakeholders on how to best respond to this content and protect Australians when online.

- Finding 5:** Misinformation typically stems from small online conspiratorial communities, but can be amplified by influential individuals, digital platform design, as well as the media.
- Finding 6:** Conspiratorial content is designed to be highly engaging, fuelling outrage, and building on a sense of community. The confluence of conspiracy theories around COVID-19 has created more paths to online misinformation.
- Finding 7:** There is some evidence of co-ordinated inauthentic activity surrounding popular misinformation narratives in Australia. Those who spread misinformation often seek to reframe global conspiratorial narratives, like QAnon, in a local context.

2.4. Impact and harms

Widespread belief in harmful misinformation can have serious impacts on individuals and society, with the potential to cause a broad range of harms. These harms can be acute, such as posing an immediate and serious threat to an individual's health and safety, or chronic, such as the gradual undermining of trust in public institutions or authoritative sources of information.

Incitements to violence

While platforms have historically acted against content posing serious and imminent threats of harms, the 2021 US Capitol riot is an example of the impact of longer-term

⁴⁸ Thomas, E. et al., [Pro-Russia vaccine politics drives new disinformation narratives](#), Australian Strategic Policy Institute, 24 August 2020.

⁴⁹ See, for example, Dziedzic, S. and Norman, J., ['Scott Morrison demands apology from China over 'repugnant' tweet showing Australian soldier threatening to kill child'](#), ABC News, 1 December 2020.

⁵⁰ Researchers, for example, have found AI generated algorithms are effective at automatically generating short and believable messages on social media containing false narratives; Knight, W. [AI Can Write Disinformation Now—and Dupe Human Readers](#), *Wired*, 24 May 2021.

chronic harms arising from the widespread belief in misinformation, and how this can spill over to the real-world as incitement to commit violent acts.

The fact that it has culminated in the storming of the Capitol, you know, it's not ridiculous – it's a real thing.

Male, 60s, focus group #6

On 6 January 2021, a mob of angry supporters of outgoing US President Donald Trump, many with links to the QAnon conspiracy, stormed the US Capitol building in a violent attack that left 5 dead and 140 capitol police officers injured. This insurrection was not spontaneous. Rather, it had been widely discussed and organised on social media, representing the culmination of the '#StopTheSteal' misinformation narrative that was first seeded online in September 2020 and escalated in the days and weeks following the November election.⁵¹

While platforms were quick to respond and quell calls for violence, Twitter's CEO later told a Congressional hearing that he takes some responsibility for allowing falsehoods about the US election to spread on Twitter in the lead up to the 6 January rally.⁵² Separately, an internal Facebook memo on the insurrection noted there were gaps in Facebook's policies around authentic coordinated activity, and noted that the company's focus on individual violations made them miss the larger harm that was occurring across the network at this time.⁵³

Despite considerable efforts to disrupt and deplatform these networks, online conspiracy theories still pose a real risk of violence. On 4 June 2021, the FBI issued a bulletin warning that some self-identified QAnon adherents could seek to engage in violence against political opponents, driven by a belief that they can no longer 'trust the plan' and have an obligation to move from 'digital soldiers' to real-world action.⁵⁴

Impacts on individual and public health

There are also many clear and significant examples of the real-world consequences of online misinformation locally, particularly in the context of the ongoing pandemic. A current challenge for Australia concerns the propagation of anti-vaccine misinformation narratives and growing vaccine hesitancy within the community. Based on the We Are Social sample, Australian misinformation conversations peaked in March 2021, driven almost entirely by growth in the anti-vaccine narrative over the previous 3 months.

A prominent 2014 study shows that exposure to anti-vaccine conspiracy theories can directly affect vaccination intentions, introducing undue suspicion about vaccine safety, while also decreasing trust in authorities.⁵⁵ Another more recent large-scale global study of vaccine sentiment expressed on Facebook pages found that those who

⁵¹ DFRLab, [#StopTheSteal: Timeline of Social Media and Extremist Activities Leading to 1/6 Insurrection](#), Just Security, February 2021.

⁵² Conger, K., [Jack Dorsey says Twitter played a role in U.S. Capitol riot](#), *The New York Times*, 25 March 2021.

⁵³ Mac, R. et al, [Facebook Stopped Employees From Reading An Internal Report About Its Role In The Insurrection. You Can Read It Here](#), BuzzFeed News, 26 April 2021.

⁵⁴ Benner, K., [The F.B.I. warns that some QAnon believers could turn to violence as predictions fail to bear fruit](#), *The New York Times*, 15 June 2021.

⁵⁵ Jolley, D. and Karen, D., [The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions](#), *PLOS One*, 9, e89177, 2014.

were undecided on vaccines tended to interact more heavily within clusters of pages sharing anti-vaccination views, compared to those with pro-vaccination views.⁵⁶

In the context of both 'data deficits' and the over-supply of information surrounding vaccine technology, trials and side effects, bad actors can take advantage of the resulting uncertainty to spread anti-vaccination misinformation to more Australians and increase levels of vaccine hesitancy.⁵⁷ Concerns about vaccine misinformation were shared by most – but not all – of our focus group participants.

But what worries me is that you might get someone old that sees that. That thinks this vaccine is not safe for me. It might put them off getting vaccinated. Because it's just these false claims that they put out and vulnerable people could just believe it, you know?

Female, 40s, focus group #12

[...] having this misinformation around the internet and so accessible is actually creating fear and danger in communities because it means that people won't get vaccinated. They don't believe what politicians, what scientists are saying, like, 'this is what we're going to do to get over this pandemic and get through it'.

Female, 18, focus group #12

It's well within our rights to be questioning it every step of the way, and we should be. It doesn't mean we're conspiracy theorists [...] I totally 100% understand and believe that COVID is a real thing [...] But sometimes I just go 'what if it's more the other way, and that the government, not that they're faking it, but that they're exercising a kind of overt control that they maybe don't need to be doing?'. [...] And suddenly, we're all having jabs and shit.

Female, 40s, focus group #7

Financial impacts

While it is difficult to quantify, misinformation also has a financial impact. A 2019 economic study estimated the annual global cost of fake news at US\$78 billion, driven by a range of factors including market volatility and losses, financial scams, public health costs, expenditure on political races, and reputational damage to brands.⁵⁸ This study, which pre-dates COVID-19, helps highlight the pervasive nature of misinformation and its growing impact across all sectors of the economy.

Locally, mobile carriers in Australia have recently incurred a range of new and unforeseen costs related to the rise and propagation of 5G and electromagnetic energy (EME) misinformation. These costs, estimated at A\$7.9 million in 2020, are discussed further in the *Impacts of anti-5G misinformation* case study below.

⁵⁶ Johnson, N. et al, [The online competition between pro- and anti-vaccination views](#), *Nature*, 582, 230–233, May 2020.

⁵⁷ A data deficit is where there are high demands for information on a topic, but credible information is in low supply – either because it doesn't exist or isn't reaching its intended audience. In a study of social media conversations, First Draft identified a range of key data deficits relating to vaccines, resulting in increasing vaccine scepticism; [Smith, R. et al, Under the surface: Covid-19 vaccine narratives, misinformation and data deficits on social media](#), First Draft, November 2020.

⁵⁸ Cheq, [The Economic Cost of Bad Actors on the Internet – Fake News](#), 2019.

Case study: Impacts of anti-5G misinformation

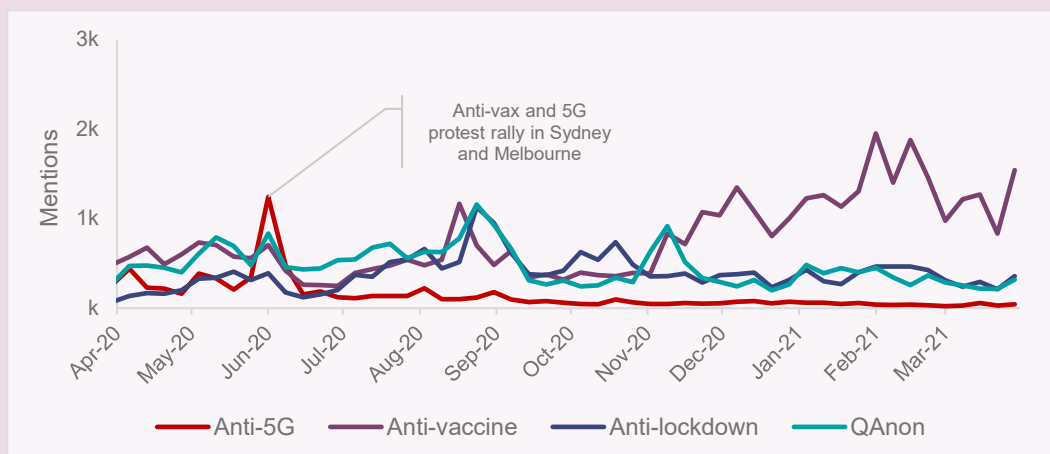
While the rollout of new wireless networks is occasionally controversial, COVID-19 has served to escalate worldwide opposition to 5G technology and deepen community concern about the negative health impacts of electromagnetic energy (EME).

From late January 2020 onwards, a growing number of online conversations appeared linking the emergence of COVID-19 to pre-existing conspiracy theories about 5G. These ranged from allegations that 5G weakened the immunity system, to claims the virus was either spread directly by 5G, was fake and covering up 5G symptoms, or had been intentionally released to hasten the 5G rollout. Some of these conversations appeared organically, while others showed evidence of bot-like behaviour or other techniques common to coordinated disinformation campaigns.⁵⁹

These conspiracies rapidly gained traction during the initial stages of lockdown, further amplified by posts from celebrities like US actor Woody Harrelson, UK boxer Amir Khan and US rapper Wiz Khalifa.⁶⁰ By early April, public concerns had escalated across Europe and the UK, resulting in reports of telecommunications engineers being threatened or harassed, and arson or vandalism attacks on at least 30 mobile towers in Britain.⁶¹ During this time, an Essential Poll showed 12% of Australians believed that the 5G network was being used to spread the COVID-19 virus⁶², and membership of Australia's largest anti-5G Facebook group surged from 6,800 to more than 48,000, fuelled by local celebrities and influencers appealing to local concerns.⁶³

Of the 4 key misinformation narratives monitored to inform this report, anti-5G had the lowest overall share of mentions over the last 12 months. As per Figure 15 below, posts with anti-5G keywords reached a peak in mid-2020 in response to small anti-5G rallies being held across Australia, but quickly dissipated relative to the other narratives. This sharp decline coincided with a range of public communications campaigns across government and industry, as well as actions taken by Facebook and other platforms to limit the reach of content linking 5G to COVID-19.

Figure 15: Volume of mentions by key misinformation narratives



Source: We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

⁵⁹ Gallagher, R. [5G Virus Conspiracy Theory Fueled by Coordinated Effort](#), *Bloomberg News*, 9 April 2020.

⁶⁰ Bruns, A., et al., [‘Corona? 5G? or both?’: the dynamics of COVID-19/5G conspiracy theories on Facebook](#), 177(1), Media International Australia, 12-29, November 2020.

⁶¹ Satariano, A. and Alba, D., [Burning Cell Towers, Out of Baseless Fear They Spread the Virus](#), *The News York Times*, 10 April 2020.

⁶² Essential Research, [Belief in Conspiracy Theories](#), 19 May 2020.

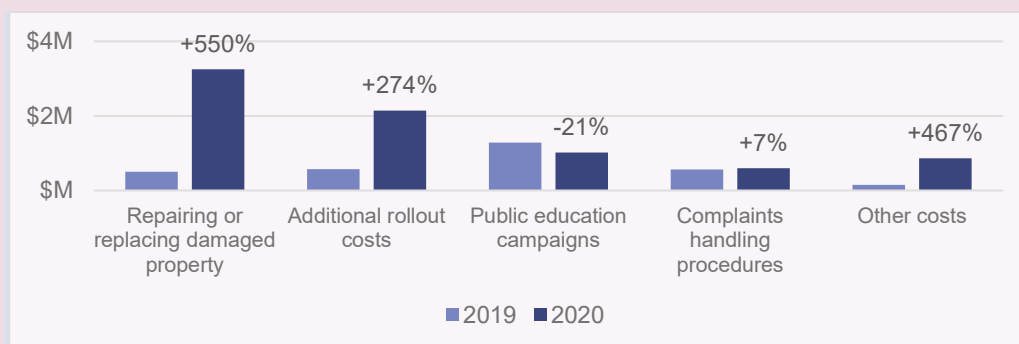
⁶³ Nicholls, S., et al. [What is the truth about 5G? Four Corners spoke to leading experts and anti-5G activists to find out](#), *ABC News*, 3 August 2020.

While there are still highly localised pockets of anti-5G sentiment – such as in Northern NSW – the research suggests that Australia’s efforts to combat 5G misinformation have been mostly successful.⁶⁴ Yet despite its relatively short-lived nature, there is a considerable financial cost associated with the spread of 5G misinformation narratives, largely borne by governments and industry, and ultimately passed on to consumers.

In late 2020, the ACMA approached the Australian Mobile Telecommunications Association (AMTA) to gauge whether the mobile industry would be interested to participate in an exercise to quantify the cost of addressing 5G and EME misinformation. Telstra, Optus, TPG Telecom and AMTA all provided the ACMA with high-level cost inputs, allowing us to estimate the financial impact across the industry.

Based on the industry data collected by the ACMA, the Australian mobile industry spent an estimated A\$11.0 million across calendar years 2019 and 2020 as a direct result of misinformation about EME and/or 5G. Most of these costs were incurred in 2020, with total annual industry expenditure more than doubling from 2019 – from A\$3.1 million to A\$7.9 million.

Figure 16: Estimated cost to the Australian mobile industry of addressing EME and 5G misinformation



Source: ACMA, based on cost estimates provided by Telstra, Optus, TPG Telecom and AMTA.

All 3 Australian carriers reported instances of arson or vandalism attacks at mobile sites that were related to the propagation of 5G/EME misinformation. While the total number of impacted sites in Australia was much lower than in other countries, replacement or repair costs at these sites nevertheless represented the largest cost category in this exercise. Carriers estimate they spent A\$3.3 million in relevant repair costs during 2020, representing a 550% increase on 2019. While not captured in the cost data, this remains an ongoing issue for some carriers, with site vandalism due to 5G/EME provocation continuing throughout 2021.

Carriers also reported a 274% increase in additional rollout costs, due to new functions like staff training to counter aggressive behaviours, and security measures at sites. Other costs in 2020 were driven by monitoring and research on misinformation, and the decommissioning and relocation of sites due to community backlash.

There are a range of larger, intangible costs associated with 5G misinformation not included in these calculations. These include the cost from delays to the rollout of 5G networks, and related reduction in productivity benefits across society. An AMTA-commissioned study from 2019 projected the mobile sector, underpinned by 5G, to be worth A\$65 billion to the Australian economy by 2023.⁶⁵ AMTA noted a portion of these productivity benefits are unlikely to be realised due to 5G misinformation.

⁶⁴ Particularly when compared to other countries, such as India; [Here's why '5G spreads Covid' is a myth, says government](#), *The Times of India*, 5 June 2021.

⁶⁵ Deloitte Access Economics, [Mobile Nation 2019: The 5G Future](#), 2019.

Finding 8: Misinformation narratives can result in a wide range of acute and chronic harms, including the erosion of trust in authoritative sources and democratic institutions over time.

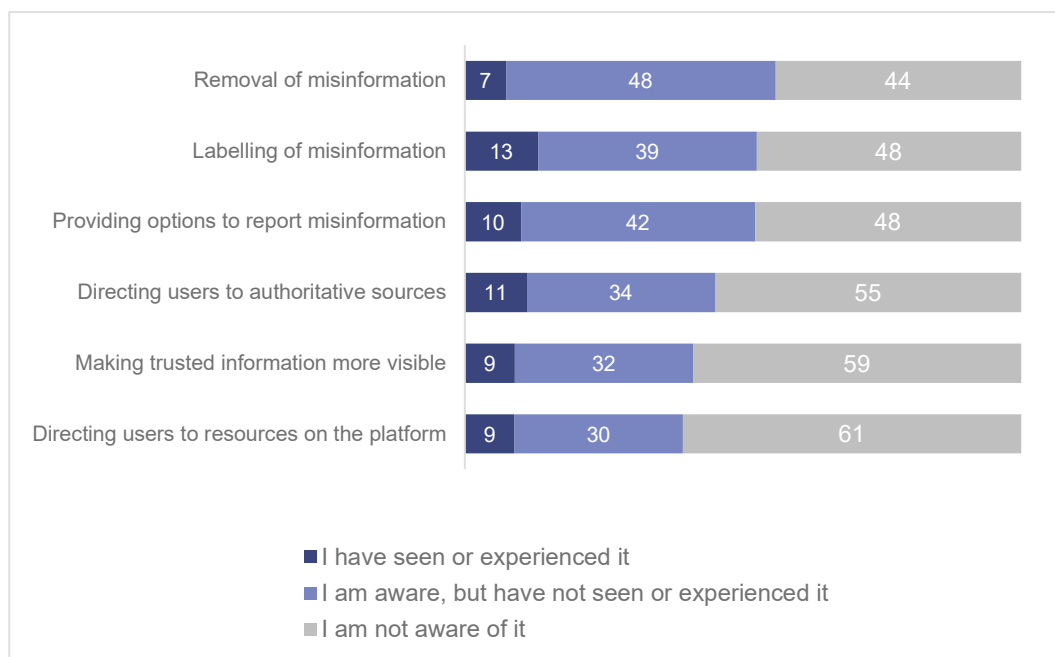
Finding 9: The real-world consequences of misinformation have been readily apparent over the past 18 months: inciting violence, undermining official health advice, and causing tangible financial impacts on governments, industry and consumers.

2.5. Platform measures and accountability

Most digital platforms have a range of existing measures to address seriously harmful, false or misleading information, and many of these policies, tools and initiatives have been strengthened over the previous 18 months in the context of COVID-19 (see Appendix C).

Ahead of the commencement of the code, N&MRC asked Australians about their knowledge of these existing efforts. There was a general awareness of some of the measures, but very few Australians reported having directly seen or experienced any of these measures (Figure 17).

Figure 17: Awareness and experience of platform measures



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

Views on platform measures

The most recognised measure – removal of content – was also the most controversial. Focus group respondents were split on the question of where platforms should draw the line. Some were of the view that all legal speech should be allowed, and only be hidden or removed if it was ‘pushing illegal activity’. Others disagreed, with one participant arguing that ‘if it’s fake, it should be off the internet’. The most common position was that there are times where it may be appropriate to remove content, but platforms should not take this step lightly given these actions place limits on speech.

These discussions were framed by media reporting on events at the time of the research, including the decision of Twitter and Facebook to ban the US President Donald Trump in January 2021 following the US Capitol riots, and Facebook’s decision to temporarily ban Australian news in February 2021, immediately before the passage of the News Media and Digital Platforms Mandatory Bargaining Code (the news media bargaining code). Participants expressed some concern about the amount of power platforms wield in making these decisions, and the potential unequal application of their policies.

I think it's a really good thing that Twitter censored [US President Donald Trump]. But then, in saying that, it's also a really interesting thing of now that they're censoring news in Australia, we're like, 'whoa whoa, you don't have that right'. It's an interesting argument to have because when it's not benefiting me, when I want to see my news on my news feed, then I'm like, well no.

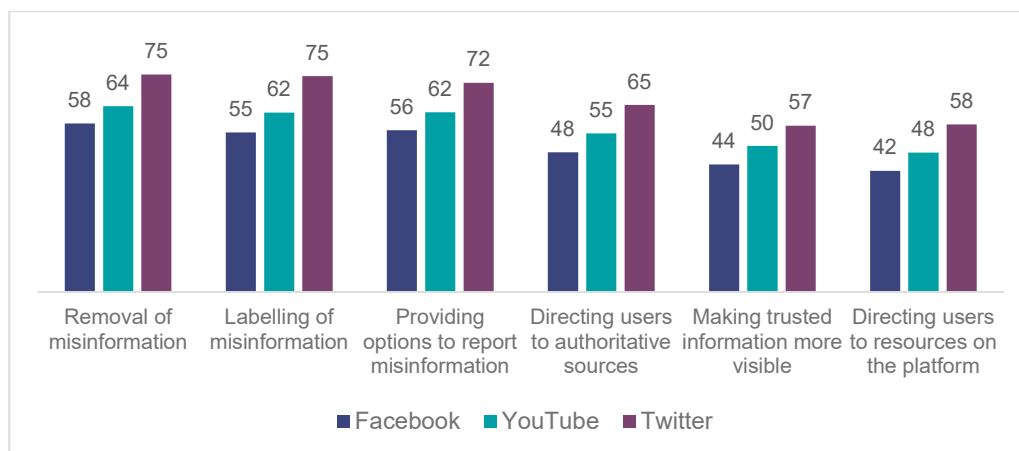
Male, 20s, focus group #1

Yeah, I think in that case, it was for the safety of people and the safety of their country. So I sort of get why they did it, but I could understand why people would think no, you shouldn't be silenced. Who makes that decision?

Female, 40s, focus group #12

There are key demographic differences in awareness of platform measures, which appear closely aligned to the characteristics of heavy users of digital platforms – those who were younger, male, and with high education were all more likely to be aware of platform measures compared to the general population.⁶⁶ Interestingly, there were also key differences between the users of particular platforms. Twitter users, for example, were considerably more likely to be aware of platform measures than Facebook users (Figure 18). This could reflect the greater publicity around Twitter’s efforts, better messaging by Twitter to its users, or demographic differences in its user base.

Figure 18: Awareness and experience of platform measures, by users of platform (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

⁶⁶ N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

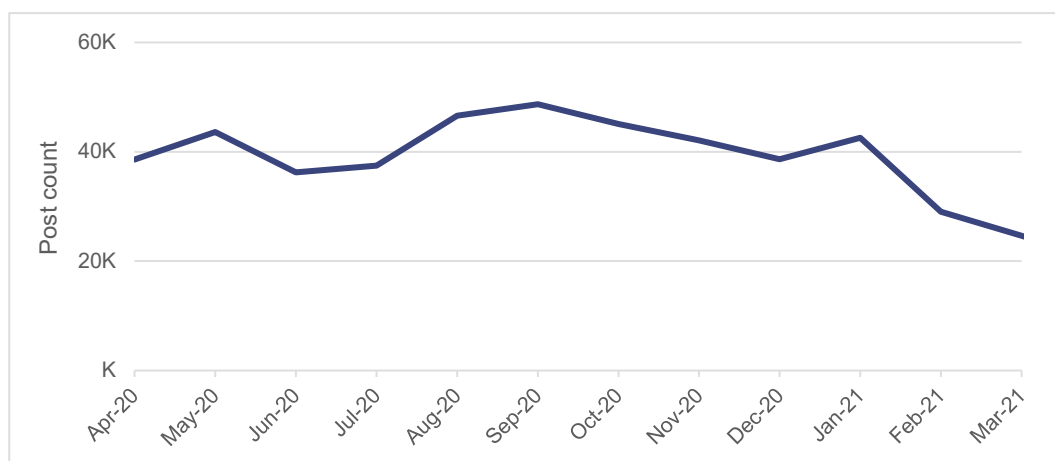
Effectiveness of platform measures

While this chapter does not seek to assess whether specific, existing measures have been effective, some high-level findings may be drawn from the research.

Although there was no noticeable decline in overall volume of misinformation conversations in the We Are Social sample over the 12-month period, conversations about certain misinformation narratives – like QAnon and anti-5G – did decline dramatically over time, in line with stronger moderation activities on these topics.

Further, by examining posts from the sample of conspiracy-driven Facebook groups over 12 months, we can see evidence of reduced amplification. The total volume of posts within these groups fell by 91% over the year ending 31 March 2021, after reaching a peak in September 2020 (Figure 19). Over the same period, there was a 192% reduction in the number of comments, 244% reduction in the number of shares, and a 302% reduction of link interactions.⁶⁷

Figure 19: Number of posts by selected conspiracy-driven Facebook groups; 1 April 2020 to 31 March 2021



Source: We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

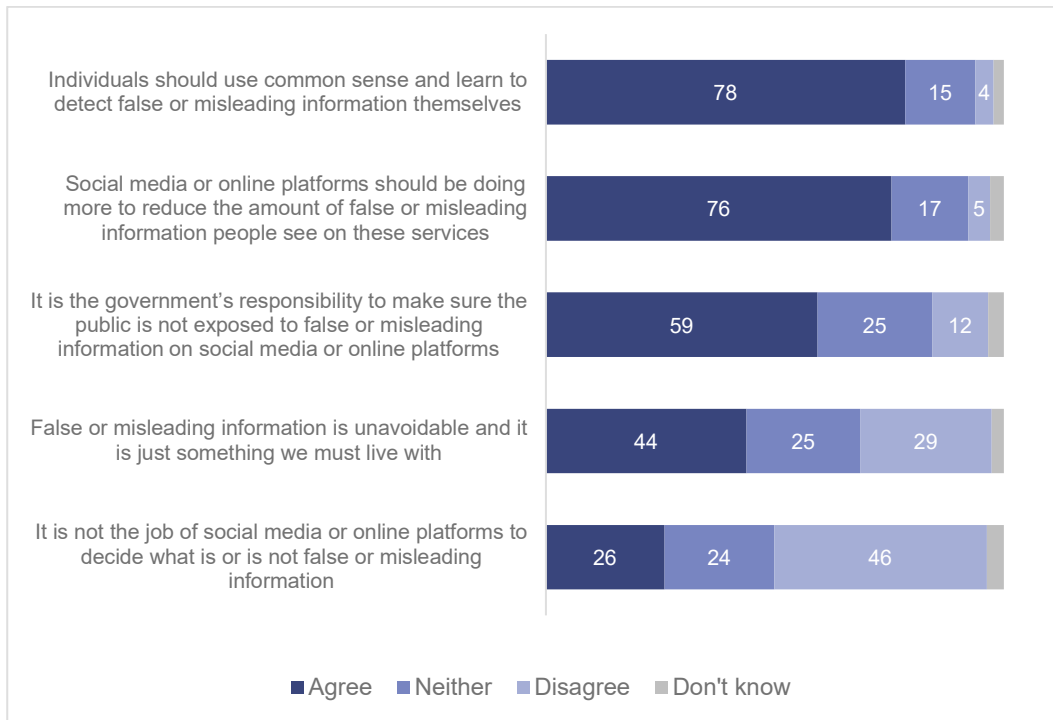
Collectively, these findings reflect a limited snapshot, and more needs to be done to better understand what measures work and how to monitor the effectiveness of platforms' moderation activities in Australia. Many researchers remain concerned that platforms are inconsistent in their approaches to detecting and addressing harmful misinformation and lack formal structures for information sharing and collaboration.

Misinformation is a shared responsibility

When asked about who bears the responsibility for addressing online misinformation, most Australians see this as an issue of shared responsibility, with 78% agreeing that 'individuals should use common sense and learn to detect false or misleading information themselves', and 76% agreeing that 'platforms should be doing more to reduce the amount of false or misleading information people see' on their services (Figure 20).

⁶⁷ We Are Social, *Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021* [unpublished].

Figure 20: Responsibility to address misinformation (%)



Source: N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

A smaller majority (59%) agreed that ‘it is the government’s responsibility to make sure the public is not exposed to false or misleading information on digital platforms’. This result may reflect a general sentiment, communicated by several participants in the focus groups, that platforms do not have sufficient commercial incentive to move voluntarily on this issue.

I think, again, in a perfect world people just wouldn't use Facebook for news but in the world we live in, I think [...] the government has to have some input into what is promoted to such a large group of people, especially because as [participant] said, Facebook doesn't exist for the people. It exists for the shareholders, it exists to generate profit, and increase its value. And I just I can't see any world where leaving it alone ends with the betterment of media and information.

Male, 20s, focus group #1

There was considerably less consensus on whether platforms should be the ones to decide on what is misinformation. Only a quarter of Australians (26%) agreed with the statement that it was not the job of social media or online platforms to decide what is or is not false or misleading information, with 46% disagreeing and saying it was. Broken down, this result differed starkly between those who were ‘informed’ about COVID-19 (15% agree) versus those who were ‘misinformed (high)’ (69% agree).

Further, while many see a role for government in addressing misinformation, focus groups were also concerned that government could be perceived by some to be untrustworthy, exercising excessive power, or impeding on individual freedoms of speech. This is an important consideration for any future regulatory reforms.

I think if it's something that has been debunked, or that is affecting a large group of the population, it will probably be the government's job to at least put that information out there. [...] I really wish that there were NGOs, and I'd love to support them, if they wanted to take this up because I don't like the government having so much power either.

Male, 20s, focus group #8

- Finding 10:** Most Australians are aware of platform measures to remove or label offending content, but few have direct experience. Early evidence suggests these steps have been somewhat effective in reducing amplification of misinformation on particular platforms.
- Finding 11:** Australians see the issue of misinformation to be one of joint responsibility – split between individual users, platforms, and government. There is some scepticism in the ability of platforms to self-regulate, and concern about government's role in regulating speech.
- Finding 12** Information on the effectiveness of platform measures is limited, and more needs to be done to better understand what measures work and to monitor the effectiveness of platform moderation activities.

3. Code development

This chapter provides an overview of the code development process and an assessment of whether it has met the expectations set out by the government.

Findings in this chapter are based on the ACMA’s observations and expertise in code development, as well as ongoing discussions with DIGI, code signatories, and other key stakeholders we have engaged with over this period.

Table 4: Timeline of key events in code development

Date	Event
December 2019	Government releases its response to the ACCC’s Digital Platforms Inquiry, calling on digital platforms to develop a code (or codes) on disinformation and news quality
January – March 2020	ACMA writes to, and meets with, most major digital platforms in Australia, seeking information on existing measures and to discuss expectations for code development
March 2020	DIGI informs the ACMA that it will take a leadership role in developing an industry-wide disinformation code and will seek involvement from both members and non-members
May 2020	DIGI provides the ACMA with a project plan. The plan signals adoption of the code by November 2020
June 2020	ACMA releases its position paper, providing advice to industry on code development, proposing a code model, and formalising its expectations for the code
October 2020	DIGI releases a draft code for public consultation. Consultation is open for 5 weeks. DIGI hosts a targeted stakeholder roundtable
February 2021	DIGI hosts a second stakeholder roundtable for submitters. DIGI publishes the final code, stakeholder submissions and a summary report, and announces the initial 6 signatories

3.1. Project timeframes

The Australian Communications and Media Authority (ACMA) will oversee the development of the code (or codes) and will report to the Government on the adequacy of the platforms' measures and the broader impacts of disinformation with the first such report due no later than June 2021.

*Government Response and Implementation Roadmap
for the Digital Platforms Inquiry, December 2019*

Position 6: The ACMA expects ...the code to be in place by no later than December 2020.

*ACMA, Misinformation and news quality on digital platforms in Australia:
A position paper to guide code development, June 2020*

In its December 2019 Implementation Roadmap, the Australian Government set an expectation that there would be a code (or codes) in force by the end of 2020. This would provide the ACMA with at least 6 months to both review the code's initial operation and develop a report to government on the adequacy of platforms' measures.

In the first quarter of 2020, the ACMA held discussions with the industry representative body DIGI as well as a range of individual digital platforms (both members and non-members of DIGI). These discussions focused on platforms' existing disinformation policies and measures, the process of developing a code, and what elements could be included in a code.

In March 2020, DIGI indicated that it would lead drafting of an industry-wide code and had engaged the University of Technology Sydney's Centre for Media Transition (CMT) and First Draft to assist in this project.

In June 2020, the ACMA released its paper *Misinformation and news quality on digital platforms in Australia: A position paper to guide code development*. This was designed to assist industry in developing a code by articulating our expectations for both the code contents and code development process, including on timelines and consultation. The paper reiterated the government's expectation that a code be in place by the end of 2020. It further proposed signatories publish individual annual action plans in January 2021 stipulating how they will meet their obligations under the code and provide a progress report to the ACMA in April 2021 to inform our report.

DIGI released a draft code for public consultation in October 2020. The final code was released publicly on 22 February 2021. At this time, DIGI announced 6 initial signatories and noted it would formally be taking on the role of code administrator.

ACMA commentary

We understand that code development can be a lengthy and difficult process, particularly for the first time. Given the complexity of the topic and the diversity of potential signatories, the work undertaken by DIGI and industry members on drafting an industry-wide code warrants recognition.

DIGI took on the responsibility of managing the project and successfully delivered on a final code that was adopted by a wide range of both DIGI member and non-member platforms. This work was undertaken in a complex environment with competing views from within and across industry, government, and civil society, and had to be completed to a relatively compressed timeframe for a first-time code.

Despite the overall achievement, it should be noted that the delay in finalisation of the code was unfortunate, albeit understandable.

It is evident that DIGI, industry members and its CMT partners understood the expected timeframe for the commencement of the code, but that development took longer than initially anticipated. DIGI has explained that its original project timeline had not accounted for the release of the ACMA position paper or the expectation that it should consult with a diverse range of potential signatories before the draft code's release. DIGI also attributed the delay to the difficulties negotiating with several non-DIGI members who had expressed interest in participating but needed more time to review.

The ACMA appreciates this was a novel and complex project and it takes time to reach consensus across an industry. We also acknowledge there were a range of competing priorities throughout 2020. These included platforms' evolving responses to the COVID-19 pandemic, and several concurrent streams of work arising from the DPI, including the introduction of the news media bargaining code.

However, it should be noted that this short delay significantly reduced the ACMA's ability to monitor and report on the effectiveness of the final code and prevented consultative work with the industry on the ACMA's proposed monitoring framework. This delay, along with the deferral of code administration matters discussed below in Chapter 4, should be recognised as a general limitation of this assessment process.

Finding 13: In leading code development, DIGI successfully managed a novel, complex and time-sensitive project, navigating a range of competing interests across a disparate group of stakeholders that included both members and non-members of DIGI.

3.2. Public consultation process

Position 6: The ACMA expects platforms to undertake an open, public consultation process when developing the code...

*ACMA, Misinformation and news quality on digital platforms in Australia:
A position paper to guide code development, June 2020*

Transparency can encourage genuine dialogue and build trust in the policy process, but in order for your consultation to be credible and effective, you need to engage with stakeholders in a way that is relevant and convenient for them. You also need to give stakeholders time to consider the information you give them and time to respond.

The Australian Government Guide to Regulation, March 2014

In the June 2020 position paper, the ACMA set out its expectation that digital platforms undertake meaningful public consultation on the drafting of the industry code. We noted that input should be sought from experts across academia, relevant government agencies, and impacted stakeholders including consumer groups and users of digital platforms.

DIGI undertook a 5-week public consultation process on their draft code, supported by a discussion paper, which was authored by CMT and provided additional background research. DIGI invited submissions from 42 organisations they had identified as relevant and ran a virtual roundtable discussion with 12 targeted academics and relevant subject matter experts.

The consultation process generated 17 public submissions from a range of stakeholders including academics, thinktanks, news media, and professional and civil society organisations. These submitters were later invited to a briefing by DIGI on how their feedback impacted the final code ahead of its release.

DIGI published these submissions on its website on 22 February 2021, coinciding with the release of the final code. DIGI also prepared and published an 18-page summary report that discussed how stakeholder feedback was addressed in the final code.



Submitter views on the consultation process

A number of submitters spoke favourably of their experiences engaging with DIGI in this process, noting they 'felt heard' by DIGI and welcomed the opportunity to discuss their views in a constructive roundtable format with subject matter experts. They observed that DIGI appeared to take the process seriously and had been responsive to stakeholder concerns, evidenced by the significant changes made between the draft and final code.

Other submitters were more critical of DIGI and questioned whether the consultation had, in fact, been genuine. They raised concerns about the limited participation in the first roundtable, and the lack of consultation on the final code, noting it was too late to share their views during the second roundtable as it was clear that nothing would change as a result. Others noted difficulties contacting or engaging with DIGI and observed that the summary report had failed to fully respond to the issues raised in their submissions.

Some feedback on the consultation process was not directly relevant and therefore not considered by the ACMA. For example, several stakeholders advocated for a mandatory code and argued that DIGI did not sufficiently respond to these calls.

ACMA commentary

DIGI ran a robust and meaningful consultation process.

It proactively identified and reached out directly to a range of relevant organisations to inform them about the consultation and provided a reasonable timeframe to accept submissions. The process attracted comments from a representative cross-section of interested stakeholders, and DIGI provided a high level of transparency over how this feedback fed into the final code.

However, there were certain aspects of the consultation that could have been improved.

Publicity of the consultation process

The ACMA expected DIGI to undertake a full and open consultation on the code. On release of the draft code, despite reaching out directly to a wide range of stakeholders, DIGI made little effort to attract a broader range of public views. It did not put out a media statement, post an update on its social media channels, or pursue any other public communication avenues to increase awareness of the consultation process beyond those organisations it had originally identified.

Given the breadth of concern regarding COVID-related misinformation in 2020, it was important that consultation on this issue be as open as possible, particularly during the code development period.

Noting that a key focus of the code is to provide more transparency, the ACMA considers more should have been done to canvass a wider range of views across the community. This includes among relevant health experts, organisations representing CALD communities, a broader cross-section of media organisations, and users of major platforms.

Limited opportunity for further engagement

Secondly, the ACMA is concerned there was a lack of opportunity for stakeholders to provide comment on the content and drafting of the final code. A recurring observation among submitters was that they would have preferred if DIGI had engaged earlier or sought additional comments on a final draft before launching the code.

This is of particular importance given the expanded scope of the final code compared to the draft code. While signatories can rightly highlight this as evidence of being responsive to stakeholder feedback, the inclusion of 'misinformation' was a significant change. One submitter told the ACMA that it knew of other organisations that would have made a submission had DIGI consulted on a code that included misinformation. Another view expressed by some submitters was that the written submission process was insufficient, and DIGI could have explored other avenues to seek feedback from a more diverse group of stakeholders, such as via public meetings.

Some of these concerns could have been mitigated if DIGI had consulted earlier and provided further opportunity to review and comment on the final draft before its finalisation. However, we recognise that DIGI was facing timing constraints and that this would have further delayed the release of the code. In these circumstances, DIGI could have instead considered holding its second roundtable session earlier, along with the earlier release of the public submissions, so that stakeholders could provide additional verbal feedback ahead of the code's finalisation.

Finding 14: DIGI undertook a meaningful public consultation process on its draft code, generating a range of feedback from academia, industry, and parts of civil society, which visibly informed the final code.

Finding 15: DIGI could have improved its consultation process with greater publicity, including promoting it through existing public communications channels and engagement with the media.

Finding 16: DIGI dealt with stakeholder feedback in a relatively open and transparent manner. However, the significant change in scope meant it would have been best practice to provide stakeholders a further opportunity to comment on the final drafting prior to finalisation.

3.3. Code signatories

Position 3: The ACMA expects that the code will cover online search engines, social media platforms and other digital content aggregation services with a major presence in Australia. The ACMA would encourage all platforms, regardless of size, to consider signing up to the code.

*ACMA, Misinformation and news quality on digital platforms in Australia:
A position paper to guide code development, June 2020*

In line with the government's DPI response, the ACMA indicated that the code should, at a minimum, cover all 'major digital platforms' (defined by the ACCC to be search engines, social media platforms or digital content aggregators with at least 1 million active monthly users).

Given the voluntary nature of the code, the ACMA also encouraged all platforms, regardless of size, to consider signing up to the code. We noted that steps to address online misinformation could also be relevant to a range of other online services that distribute news and information to Australians, including smart devices, online forums, podcast aggregators and closed group messaging services.

Facebook, Twitter, Google, Microsoft, TikTok and Redbubble were named as initial code signatories on release of the final code on 22 February 2021. In May 2021, DIGI announced that Adobe and Apple had also signed up to the code.

Table 5: Code signatories

Signatory	Covered service(s) ⁶⁸	Start	DIGI member	EU code signatory
Facebook	Facebook Instagram	22 Feb 2021	✓	✓
Google	Google Search Google Ads YouTube	22 Feb 2021	✓	✓
Microsoft	Company-wide initiatives for its consumer-facing products (including Bing and LinkedIn)	22 Feb 2021		✓
Redbubble	Redbubble	22 Feb 2021	✓	
TikTok	TikTok	22 Feb 2021		✓
Twitter	Twitter	22 Feb 2021	✓	✓
Apple	Apple News	25 Feb 2021		
Adobe	Content Authenticity Initiative	20 Apr 2021		

The code does not apply to all products and services of signatories. As discussed in Chapter 4, it is limited to those that deliver, to end users in Australia, user-generated content (including sponsored and shared content) and/or content that is returned and ranked by Search Engines in response to user queries. The code does not ordinarily apply to news content, private messaging services, email, or enterprise services.

ACMA commentary

Based on publicly available reporting on the number of active monthly users, and our assessment of web traffic, app store rankings and survey data, it is evident that the code covers almost all major digital platforms in Australia.

We particularly welcome the inclusion of platforms that are not existing DIGI members (Microsoft, TikTok, Apple and Adobe), and platforms that are not signatories to the European Union (EU) Code of Practice on Disinformation (Apple, Adobe, and Redbubble). This reflects the efforts of DIGI in seeking broad code coverage, and

⁶⁸ Note these services are drawn from signatory opt-in nomination forms and initial annual reports, [published online](#) by DIGI on 22 May 2021.

signals strong commitment across the entire industry to voluntarily address the issue of online disinformation and misinformation in Australia. All code signatories should be commended for their decision to participate in this scheme.

However, the ACMA remains concerned about the lack of transparency surrounding the signing of new digital platforms after the code's commencement date. Apple, in particular, signed up to the code on 25 February 2021, 3 days after the code came into force. Disappointingly, Apple did not publicly acknowledge its involvement until DIGI released the initial signatory reports 3 months later, on 22 May 2021. DIGI should ensure that any future signatories to the code are prepared to announce their participation as soon as practicable after signing.

Non-signatories

Despite the broad coverage of the code, there are a small number of popular social media or search engine services that have not participated in this process. This may be due to a lack of interest or perceived relevance, limited regulatory capacity or expertise in the region, or a lack of awareness of the process. Table 6 provides a list of the most popular non-signatory services by a variety of usage metrics.

Table 6: Usage metrics of key services not covered by the code

Digital platform service	Usage in past month ⁶⁹	Usage in past week ⁷⁰	Unique monthly website visitors (excl. apps) ⁷¹ (million)	Total monthly website visits (excl. apps) ⁷² (million)	Top Android apps by usage rank ⁷³
Snapchat	29%	16%	0.5	0.6	4
Pinterest	26%	11%	6.1	9.5	85
Reddit	19%	9%	6.9	58.6	44
Twitch	13%	-	1.1	17.0	156
Tumblr	10%	-	1.4	8.2	-
Discord	-	-	1.1	8.5	39
DuckDuckGo	-	-	0.9	23.8	166

Note: Data on unique monthly website visitors is not comparable to monthly active users and does not capture usage of mobile apps. Survey data not available for all platforms.

⁶⁹ Based on the GlobalWebIndex Q3 2020 survey of approximately 16,000 Australian internet users aged between 16-64; We Are Social, [Digital 2021: Australia](#), January 2021.

⁷⁰ Based on a survey of 2,659 adult Australians between December 2020 and January 2021; N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

⁷¹ Sum of all unique website visits, from Australia, during March 2021. Includes domain and all meaningful subdomains, across both desktop and mobile web; SimilarWeb, [Top Websites – Custom Industry/Unique Visitors/Australia](#).

⁷² Sum of all non-unique website visits, from Australia, during March 2021. Includes domain and all meaningful subdomains, across both desktop and mobile web; SimilarWeb, [Top Websites – Custom Industry/Monthly Visits/Australia](#).

⁷³ 'Usage rank' is calculated by a SimilarWeb algorithm that factors in 'Current Installs' and 'Active Users', providing a ranking of the top 1,000 free Android apps for the last 28 days. Comparable rankings are not available for Apple iOS apps; SimilarWeb, [Top Apps – Google Store/Top Free/Australia](#), 30 April 2021.

The most notable non-signatory is Snapchat. Snapchat is one of the most popular mobile apps in Australia, with an estimated 6.4 million active monthly users⁷⁴, and a high app usage rank. Against these measures, Snapchat appears to easily meet the threshold of a 'major' platform and should be encouraged to participate in the code.

Pinterest and Reddit could also be considered major platforms by some metrics. Both have high volumes of web-based traffic, each boasting over 6 million unique monthly visits. Of the 2, Reddit is more commonly used as a source of general news and information, and based on the N&MRC survey, is more widely regarded as a potential source of misinformation.⁷⁵

While having less than 1% market share, privacy-oriented search engine DuckDuckGo is the third most popular search engine in Australia, has experienced considerable growth over the last 12 months,⁷⁶ and should also be encouraged to participate in the code. Research out of the UK suggests that those who get 'a great deal' or 'a fair amount' of COVID-19 information from DuckDuckGo have much higher levels of vaccine hesitancy than the general population, and are also much more likely to believe that reporters, scientists and government officials are involved in a conspiracy to cover up important information about the coronavirus.⁷⁷

It should be reiterated that private messaging platforms remain outside the scope of the code. As such, popular digital platform services like Facebook Messenger, WhatsApp and WeChat are not covered, despite growing concerns that these platforms are potential hotspots for misinformation. As discussed in Chapter 2, WeChat has low overall usage within the community but high perceived volumes of COVID-19 misinformation among its user base – second only to that of Facebook.⁷⁸

Alternative social media services

As also discussed in Chapter 2, smaller, encrypted private messaging, message board and alternative social media platforms like Signal, Telegram, 4chan, 8kun, Gab and Parler are also clear vectors for disinformation and misinformation content. These services promote themselves as bastions of free speech and have minimal or less restrictive content moderation policies, attracting conspiratorial communities that may no longer be accepted on mainstream platforms. In particular, the use of Telegram and Signal in Australia has increased in recent months, largely driven by concern over WhatsApp policy changes and increased content moderation on other platforms.⁷⁹

Use of these services in Australia is currently too low for any to be considered a 'major platform', and their respective stances on moderation make it highly unlikely that any would consider signing up to a voluntary code of practice in Australia.

Nevertheless, governments, researchers and civil society should continue to actively engage with these platforms on harmful misinformation narratives, and industry should consider options for how to best engage with them as part of ongoing industry-wide efforts to address disinformation and misinformation.

⁷⁴ SocialMediaNews, [Social Media Statistics Australia – April 2021](#), published 1 May 2021.

⁷⁵ N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

⁷⁶ StatCounter, [Search Engine Market Share Australia](#), May 2021.

⁷⁷ University of Bristol and King's College London, [Coronavirus conspiracies and views of vaccination](#), 31 January 2021.

⁷⁸ N&MRC, COVID-19: Australian News & Misinformation Longitudinal Study, 2021 [unpublished].

⁷⁹ We Are Social, Social media insights into how online misinformation and disinformation are being spread across social platforms in Australia, May 2021 [unpublished].

Finding 17: The bulk of 'major platforms' in Australia have signed up to the code. As such, it should be regarded as an industry-wide initiative.

Finding 18: DIGI should continue to encourage other popular platforms, like Snapchat and Reddit, to sign up to the code, even if they do not meet the proposed threshold of 1 million active monthly users. DIGI should actively publicise the involvement of any additional code signatories as soon as practicable after their signing.

Finding 19: Industry participants should consider the role of private messaging platforms and smaller alternative platforms in the amplification of disinformation and misinformation and explore options for how these platforms could be included within the code framework.

4. Assessment of the code

This chapter provides an overview of the code and an assessment of whether, in the ACMA’s view, it has met the expectations set out by the government. Findings in this chapter draw on submissions to the public consultation process, discussions with DIGI, code signatories and other key stakeholders, and the ACMA’s expertise in code development.

4.1. Regulatory approach and framework

The code takes an outcomes-based regulatory approach, specifying objectives and outcomes that signatories commit to achieve. This section examines the general approach taken in the code.

Table 7: Code objectives and outcomes

Objectives	Outcomes
Objective 1: Provide safeguards against Harms that may arise from Disinformation and Misinformation.	<p>Outcome 1a: Signatories contribute to reducing the risk of Harms that may arise from the propagation of Disinformation and Misinformation on digital platforms by adopting a range of scalable measures.</p> <p>Outcome 1b: Users will be informed about the types of behaviours and types of content that will be prohibited and/or managed by Signatories under this Code.</p> <p>Outcome 1c: Users can report content or behaviours to Signatories that violates their policies ... through publicly available and accessible reporting tools.</p> <p>Outcome 1d: Users will be able to access general information about Signatories’ actions in response to reports.</p>
Objective 2: Disrupt advertising and monetisation incentives for Disinformation.	Outcome 2: Advertising and/or monetisation incentives for Disinformation are reduced.
Objective 3: Work to ensure the integrity and security of services and products delivered by digital platforms.	Outcome 3: The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.
Objective 4: Empower consumers to make better informed choices of digital content.	Outcome 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.
Objective 5: Improve public awareness of the source of Political Advertising carried on digital platforms.	Outcome 5: Users are better informed about the source of Political Advertising.
Objective 6: Strengthen public understanding of Disinformation and Misinformation through support of strategic research.	Outcome 6: Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.
Objective 7: Signatories publicise the measures they take to combat Disinformation and Misinformation.	Outcome 7: The public can access information about the measures Signatories have taken to combat Disinformation and Misinformation.

ACMA commentary

It is extremely positive to see industry, steered by DIGI, come together to develop a single code of practice. A single code should promote a consistent approach to dealing with misinformation across platforms, while providing efficiencies through standardised administration, complaints handling, and reporting processes. It also means users only need to go to a single place to understand the protections offered by signatories. This should promote confidence in industry to manage the range of harms associated with misinformation.

The code is framed to address the Australian environment. While Australia or Australians are not directly referenced in the code's objectives and outcomes, the definitions limit the scope of the code to digital content targeted at Australian users.⁸⁰ The effectiveness of the code in this area will depend on the extent to which signatories tailor their measures for Australia, and provide relevant, local information and data in their annual reporting on issues and measures affecting Australian users.

On issues of language, the code incorporates complex definitions and uses technical jargon that may make aspects of the code unclear to users and the general public. The development of definitions has been challenging due to the relatively novel nature of the problem and lack of consensus on definitions among industry, researchers and international organisations.⁸¹ However, the ACMA considers that clear definitions and simple language would make the code more accessible to the public and increase transparency of platform measures.

The following provides a more detailed discussion on specific issues relating to the regulatory approach and framework adopted by the code, including the guiding principles, outcomes-based approach, and opt-in framework.

Guiding principles – protection of users' rights

The code opens with a preamble that contextualises the problem and explains to readers why the code was established. It also includes 7 guiding principles, which are designed to inform the operation of the code and assist signatories in developing suitable measures that do not, for example, impede on existing user protections or the security of their services.

A key concern for government, articulated in its response to the DPI, is that the code balances any interventions with rights to freedom of expression and speech. The code acknowledges this issue upfront, with its first guiding principle focused on protection of freedom of expression:

Protection of freedom of expression: Digital platforms provide a vital avenue for the open exchange of opinion, speech, information, research and debate and conversation as well as creative and other expression across the Australian community. Signatories should not be compelled by Governments or other parties to remove content solely on the basis of its alleged falsity if the content would not otherwise be unlawful.⁸²

This guiding principle also encourages signatories to be cognisant of the need to protect internationally recognised human rights in developing proportionate responses to disinformation and misinformation, including, but not limited to, freedom of speech. Similar statements are also included in the EU Code (Appendix D).

⁸⁰ Code provision 3.1.

⁸¹ Code provision 1.2.

⁸² Code provision 2.1.

The ACMA considers the preamble and statement of guiding principles a welcome inclusion, particularly in flagging the need for signatories to protect user privacy, support independent researchers, and balance protection from harms caused by the propagation of disinformation and misinformation with freedom of expression and other rights. It is important that no obligation is placed on signatories to remove content merely on the basis of inaccuracy, where a threshold of serious harm is not met.

However, platforms still retain a responsibility to improve the online information environment by implementing appropriate measures where a threshold of serious harm is not met. Examples could include labelling of fact-checked content, introducing friction to counter virality, or increasing transparency of information sources.

Outcomes-based approach

The ACMA considers the code's outcomes-based approach is well suited to the nature of the problem and the disparate business models and services of the major digital platforms. Overall, the objectives and outcomes of the code are framed to provide protections for both users and the general public. This is a key success of the code.

Under an outcomes approach, entities have the flexibility to develop their own measures in a way that best reflects their services and business models. This allows for innovation as industry adapts to a dynamic problem and makes progress towards the achievement of common objectives.

The outcomes approach also provides platforms with the flexibility to implement measures to counter disinformation and misinformation in proportion to the risk of potential harm. This allows platforms to balance the need to address potential harms with freedom of expression and other rights. Section 6 of the code details the need for proportionality and the criteria that platforms may use when assessing the appropriateness of their measures. Measures such as the removal of content or user accounts can be appropriately limited to situations where there is a very high risk of harm, while less stringent measures may be applied to lower-risk content in accordance with platforms' policies and procedures.

To be successful, an outcomes-based code depends on a high degree of commitment from industry to work towards common, measurable outcomes established by the code, and to demonstrate this through clear and robust performance reporting. More information on an outcomes-based approach can be found in the ACMA's position paper.

Opt-in framework

A key feature of the code is its opt-in framework. All signatories must commit to outcome 1(a) – reducing the risk of harms arising from disinformation and misinformation – and to publish an annual report (see Table 8). Signatories are then free to opt-in to the other objectives and outcomes. This approach is designed to accommodate digital platforms that 'operate vastly different businesses which offer a wide and constantly evolving variety of services and products'⁸³ and 'the need of the Signatories to choose those measures which are most suitable to address instances of Disinformation and Misinformation' on their services.⁸⁴ DIGI has noted that some signatories may not have signed up to the code without the flexibility provided by the opt-in model.

As noted above, the code has only 2 mandatory commitments. This provides a low minimum standard for collective industry action, and by not specifying criteria for

⁸³ Code provision 1.1.

⁸⁴ See code provision 1.5.

opting out, potentially reduces transparency. Several submissions on the draft code made similar criticisms of the opt-in model.⁸⁵

The code's effectiveness will be dependent on widespread implementation of measures to address disinformation and misinformation wherever such content exists. While signatories' reports show that they have by and large committed to all outcomes relevant to their services (see Chapter 5), the ACMA considers that the code should be strengthened by implementing an opt-out framework. Under such a framework, platforms would be permitted to opt out of an outcome only where that outcome is not relevant to their services. The recent European Commission (EC) guidance on the EU Code suggests a similar approach (see Appendix D for broader discussion on international approaches).⁸⁶

When opting out, platforms should be required to provide justification demonstrating that the outcome is not relevant to their services. The ACMA considers that this approach would provide signatories with sufficient flexibility to opt out where a service is clearly not relevant, while providing greater transparency about signatories' approaches and encouraging industry action. In addition, the proportionality principle set out in Provision 6.1 means that a platform's size, nature and available resources should be taken into account when considering the appropriateness of a platform's measures, their ability to contribute to research, and initiatives under Outcomes 6 and 7 respectively.

As well as a high degree of industry commitment to take action to achieve agreed outcomes, an outcomes-based model requires:

- > concrete, measurable outcomes and a clearly defined scope that identifies and directly targets the problem
- > comprehensive reporting against agreed key performance indicators that measure progress towards code outcomes
- > a robust code-administration framework supported by a consistent program of evaluation and review.

The code is assessed against these requirements in the following sections.

⁸⁵ The Australian Associated Press noted the need for commitment to common objectives, stating that it is 'essential that the common purpose is clearly identified in Australia's code, and accepted by all signatories' and that 'expectations around engaging with the code to the fullest extent possible should be clearly stated'. The ABC and SBS submissions observed that the opt-in model may reduce the incentive for platforms to expand or improve their current initiatives. Reset Australia expressed scepticism that the opt-in model would be effective in tackling disinformation. The Australian Muslim Advocacy Network argued that an opt-out model would provide more scrutiny, with applications to opt out of an outcome to be made to an independent administrator.

⁸⁶ European Commission, European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021, p. 6.

Finding 20: DIGI has developed an outcomes-based code that has allowed platforms with a range of business models to sign up to a single code.

Finding 21: The code objectives and principles meet the government objective of striking a balance between encouraging platform interventions and protecting freedom of expression, privacy and other rights.

Finding 22: The code should be strengthened by taking an opt-out approach. Opting out of an outcome should be permitted only where the outcome is not relevant to the signatory's services. Signatories should provide adequate justification when opting out.

4.2. Code scope

The scope of the code hinges on the definitions of disinformation (and inauthentic behaviour), misinformation and harm. These definitions are outlined below.

Disinformation

The aspect of Disinformation that this Code focuses on is:

- A. Digital Content that is verifiably false or misleading or deceptive;
- B. is propagated amongst users of digital platforms via Inauthentic Behaviours; and
- C. the dissemination of which is reasonably likely to cause Harm. [3.2]

Inauthentic Behaviour

Inauthentic behaviour includes spam and other forms of deceptive, manipulative or bulk, aggressive behaviours (which may be perpetrated via automated systems) and includes behaviours which are intended to artificially influence users' online conversations and/or to encourage users of digital platforms to propagate Digital Content. [3.5]

Misinformation

Misinformation means:

- A. Digital Content (often legal) that is verifiably false or misleading or deceptive
- B. is propagated by users of digital platforms; and
- C. the dissemination of which is reasonably likely (but may not be clearly intended to) cause Harm. [3.6]

Harm

Harm means harms which pose an imminent and serious threat to:

- A. democratic political and policymaking processes such as voter fraud, voter interference, voting misinformation; or
- B. public goods such as the protection of citizens' health, protection of marginalised or vulnerable groups, public safety and security or the environment. [3.4]

One of the key strengths of the code is that it covers both disinformation and misinformation. This is a significant improvement on the consultation draft and provides Australian users with stronger protections than those afforded under the

equivalent EU Code.⁸⁷ DIGI and code signatories should be applauded for listening to stakeholder feedback on this issue and broadening the scope of the code.

The ACMA also welcomes the inclusion of a harm threshold. This acts as a fundamental safeguard against platforms unnecessarily impinging on freedom of expression. We are, however, concerned that the effectiveness of the code will be limited by an excessively narrow definition, or interpretation, of harm.

The code does not require signatories to take action against content unless it is reasonably likely to result in ‘serious’ and ‘imminent’ harm. The ACMA agrees that a threshold of ‘serious’ harm is appropriate. This is in line with the DPI which recommended a high threshold of ‘serious public detriment’ to avoid undue limits on free expression.⁸⁸

However, the requirement that harm must also be ‘imminent’ introduces a temporal element, which may be interpreted differently by signatories. If read narrowly, the ‘imminent’ test would likely exclude a range of chronic harms that can result from the cumulative effect of misinformation over time, such as reductions in community cohesion and a lessening of trust in public institutions.⁸⁹ As outlined in Chapter 2, these types of chronic harms can increase vaccine hesitancy, promote disengagement from democratic processes, and result in a range of tangible, real-world harms to both individual users and society at large.

As it could have the effect of potentially excluding from the code a significant amount of material of concern, the ACMA considers that the term ‘imminent’ should be removed from the harm definition.

Services covered by the code

Section 4 of the code specifies the types of services and products that are covered by the code.

Provision 4.1 articulates the scope of the code as applying to ‘services and products that deliver to end users in Australia:

- A. user-generated (including sponsored and shared) content; and/or
- B. content that is returned and ranked by Search Engines in response to user queries.

Provision 4.2 explicitly excludes the following services:

- A. private messaging services including those provided via software applications;
- B. email services including those provided via software applications;
- C. Enterprise Services.⁹⁰

⁸⁷ The EC guidance on the EU Code recommends that the code should be strengthened to commit signatories to have in place policies and measures to mitigate the risks posed by misinformation where there is a significant public harm dimension. European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021, p. 5.

⁸⁸ Noting the DPI recommendation was in the context of a mandatory code addressing complaints about content, where compliance would be assessed against whether code rules are followed in particular cases.

⁸⁹ See ACMA, [Misinformation and News Quality on Digital Platforms in Australia](#), p.12.

⁹⁰ The code defines Enterprise Services as software and services including cloud storage and content delivery services which are designed for the use of a specific organisation.

Provision 4.3 notes that the list of excluded services and products is not intended to be exhaustive as new services and products are likely to emerge, some of which will not be relevant to the code.

The code does not require signatories to nominate which of its services are covered. Noting the open-ended nature of Provision 4.3, the ACMA considers that signatories should be required to report annually on which services and products are covered by the code. This is discussed further in Chapter 5.

Private and group messaging services

Under Provision 4.2(A), private messaging services are explicitly excluded from the code. However, the code does not provide a definition of private messaging, which creates some uncertainty about the range of services intended to be captured by this exclusion.

Most private messaging services offer group message functionality, which can allow for conversations between hundreds or even thousands of users at once. Some of these larger groups take on the characteristics of a semi-public community or channel, extending beyond friends and family to include strangers with shared interests. Our environmental assessment in Chapter 2 indicates that the propagation of disinformation and misinformation on these channels is increasingly problematic, particularly on smaller, alternative platforms that allow for very large group sizes.⁹¹

However, discussions with DIGI indicate that the intention is to exclude all messaging products and services, including those that allow messaging to large groups. This is justified on both privacy grounds, and due to technical limitations, such as end-to-end encryption.

The ACMA agrees that private messaging services should be treated differently from other digital platform services, and that there should be no requirement or expectation that digital platforms should monitor private conversations between users. There are, however, steps that platforms can – and in some cases, do – take to reduce the risk of misinformation on messaging services, without needing to access or view the content. Many of these measures are designed to reduce the amplification of content, for example by introducing ‘friction’ that limits the speed or reach of forwarded messages or shared links. As these services fall outside the scope of the code, platforms are not required to be transparent about the nature and extent of the problem on messaging services or about the effectiveness of their measures.

The ACMA’s position paper recommended that, given their role in spreading and amplifying misinformation, online groups and semi-public channels such as one-to-many or many-to-many messaging services should be covered by the code.

Stakeholders have also expressed concern about the exclusion of private messaging services. In its submission to the draft code, the ABC noted the lack of clarity in this exclusion, given the lack of a definition. The Digital Media Research Centre at Queensland University of Technology suggested that private messaging be included given the significant amount of disinformation spread through instant messaging. The EC has also recently recommended the expansion of the EU Code to cover private messaging services.⁹²

The ACMA considers there would be substantial public benefit in including messaging services within the scope of the code, with the caveat that this should not entail any

⁹¹ Telegram, for example, allows messaging to groups of 200,000, and Signal to groups of 1,000.

⁹² European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021, p. 5.

obligation on platforms to monitor or censor the content of private messages. As private, closed and semi-public groups or pages on social media platforms are already included within the scope of user-generated content,⁹³ the inclusion of one-to-many messaging services would support greater consistency across service types.

News aggregation services

It is also not clear whether news aggregation services are covered by the code. Both the DPI and the ACMA's position paper proposed that major platforms providing digital content aggregation services should be included. Provision 4.1 appears to exclude these services as they neither involve user-generated content nor are considered search engines. DIGI has advised, however, that it considers these services to be covered by the code, and Apple has signed up to the code on the basis of its Apple News aggregation service.

News aggregators play a key role in disseminating news and information through the online information ecosystem. Many stakeholders have noted the importance of quality news being readily available online. If the code is to offer a comprehensive, industry-wide approach to misinformation, disinformation and news quality, it is vital that news aggregators delivering content to Australians are explicitly included within the scope of the code. This should be clarified in later code revisions.

Content covered by the code

Section 4 of the code also places limitations on the types of content that is to be covered by the code.

Provision 4.4 excludes the following types of content from the operation of the code, unless signatories determine that specific instances of its propagation clearly fall within the scope of disinformation:

- A. content produced in good faith for entertainment (including satire and parody) or for educational purposes;
- B. content that is authorised by an Australian State or Federal Government;
- C. subject to sections 5.21 to 5.23, Political Advertising or content authorised by a political party registered under Australian law; and
- D. news content that is the subject of a published editorial code which sets out content standards and or/complaints mechanisms.

Provision 4.5 states that signatories may, at their discretion, 'implement policies and procedures which govern the dissemination by users on their platforms of the types of content excluded from the operation of the provisions of the code under section 4.4, where signatories determine such content is reasonably likely to cause Harm.'

The ACMA considers the exclusion of entertainment and educational content produced in good faith, government-approved content, and authorised political content is appropriate given the need to balance interventions with freedom of expression. On issues of professional news content and political advertising, however, greater clarity would be welcomed.

Professional news content

Provision 4.4.D of the code excludes news subject to published editorial standards, except in clear and specific instances of disinformation.⁹⁴ Although this wording seems to exclude professional news from the code as a whole, DIGI has advised that this is not the intention. Instead, the purpose is to clarify that platforms will not consider

⁹³ Code provision 4.1.A.

⁹⁴ Code provision 4.4.D.

professional news as misinformation under the code. That is, the code places no obligation on platforms to extend counter-misinformation measures to professional news content.⁹⁵

The ACMA acknowledges that professional news should be treated differently from other types of online content, as most news content is already covered by separate regulatory frameworks that promote accuracy and impartiality in reporting, and provide separate avenues for complaints handling. However, there is concern that news content that does not present a high risk of harm at the publisher level can sometimes present a higher risk once it is taken out of that context.⁹⁶ For example, Chapter 2 outlines how links to news articles from reputable Australian media outlets are commonly shared within conspiratorial communities to support misinformation narratives. The code also leaves open the question of how news from sources not subject to an Australian professional editorial code, such as non-Australian sources, is to be treated.

In the ACMA's view, the exclusion of professional news from misinformation is unnecessary. The outcomes-based model can accommodate the different treatment of professional news compared to other content, allowing platforms to apply different measures in consideration of the editorial standards and complaints processes that may already apply.

Political and issues-based advertising

The ACMA welcomes the objective to improve public awareness of the source of political advertising (Objective 5). Political advertising is otherwise excluded from the code, except where specific instances of it clearly fall within the scope of disinformation. This refers to content that is propagated by inauthentic behaviours, such as spam, bots, fake accounts or deceptive behaviours including foreign interference and other information operations.

The ACMA considers this exclusion to be appropriate. Political advertising is covered by existing electoral law obligations, and the protection of freedom of political expression is critical.

However, it is unclear whether issues-based advertising is also excluded from the code.⁹⁷ Issues-based advertising includes sponsored and paid-for content that is intended to bring awareness to, advocate for, or call for action on certain topics that are widely discussed in the public sphere, such as political and social issues.

Issues-based advertising is a known vector of misinformation. There is particular concern about the ability of micro-targeting technologies, which rely on user data, to direct advertisements containing false or misleading information at particular groups and not others. Micro-targeting can reduce transparency, as advertisements are visible only to those who are targeted.

To alleviate these concerns, the code should include a clear definition of issues-based advertising and the scope of Objective 5 should be extended to include issues-based advertising. Providing users with greater transparency about the source of issues-

⁹⁵ Email from DIGI 'Additional ACMA questions', received 15 June 2021.

⁹⁶ For example, corrections made by the publisher may not be clear to users once the original article has been shared; publishers no longer have control over how the content is used; the content may be manipulated; or removal of original context may make the content misleading.

⁹⁷ The definition of political advertising in the code includes advertisements that 'advocate for the outcome of a political campaign concerning a social issue of public concern in Australia.' The meaning of 'advocate for the outcome of a political campaign' is not defined.

based advertising would increase users' awareness of why they are being targeted and inform their decision-making about important social issues.

Several platforms currently have separate policies on issues-based advertising. Explicitly including issues-based advertising within the scope of the code would increase transparency of their measures and encourage other platforms to implement similar measures.

There are concerns that placing limits on issues-based advertising would unduly limit political expression. Acknowledging this, the EU code commits signatories to improving the transparency of issue-based advertising, and tasks them with the development of a working definition of issue-based advertising which does not limit reporting on political discussion and the publishing of political opinion. The ACMA recommends a similar approach be taken in the Australian code.

Paid and sponsored content

The ACMA considers that platforms have a greater level of responsibility for paid and sponsored content, given their greater control over the content and the monetary benefit they receive. The potential to target content at individuals or groups also increases the potential for harm caused by disinformation and misinformation propagated through advertising channels. Signatories should take a more active role, and implement more proactive measures, in relation to this type of content.

The treatment of paid and sponsored content should be made clearer in the code. While sponsored content is explicitly included within the scope of user-generated content, the term is not defined in the code. Improved clarity around the operation of these arrangements would be beneficial for platforms, advertisers, and users.

Finding 23: The code covers both disinformation and misinformation. This is one of the key strengths of the code, and is an improvement on the current EU Code.

Finding 24: The definition of harm in the code is too narrow to provide adequate safeguards against the full range of harms caused by the propagation of disinformation and misinformation.

Finding 25: Private messaging services should be included within the scope of the code as these are known vectors of disinformation and misinformation. These should be included with appropriate caveats on the right to privacy.

Finding 26: The code should clarify that the exclusion of professional news content applies only to the application of counter-misinformation measures. It should also clarify that news aggregation services are in scope.

Finding 27: The treatment of paid and sponsored content should be clearer in the code. This should include a clear definition of sponsored content.

Finding 28: In addition to improving public awareness of the source of political advertising, the code should also cover the source of issues-based advertising.

4.3. Code objectives, outcomes and commitments

There is room for improvement in the drafting of the code. Some outcomes could be considered outputs, and this may encourage reporting focused on providing data to quantify measures implemented by platforms, rather than data on the effectiveness of those measures. Developing and embedding metrics in the reporting framework currently under development may address these concerns.

A positive element in the code is the inclusion of example measures under each code outcome covering a wide range of approaches. However, the code would benefit from more detailed commitments that lay a foundation for the development of consistent, evidence-based measures across the industry. Examples are discussed under the relevant sections below.

Safeguards against disinformation and misinformation

The list of example measures at 5.9 includes a wide range of both proactive and enabling measures to reduce the impact of disinformation and misinformation. Importantly, it includes the adjustment of ranking algorithms to reduce user exposure.

More-detailed, structured commitments to address disinformation and misinformation might include:

- > consideration of a common framework for the assessment of harm that can help structure platforms' internal decision-making
- > the establishment of mechanisms for the exchange of information between platforms on emerging disinformation and misinformation risks
- > concrete commitments to mitigate risks of recommender systems fuelling the viral spread of disinformation, as recently recommended in the EC's guidance for strengthening the EU code.⁹⁸

Disrupting advertising and monetisation incentives for disinformation

The provisions under Objective 2 are focused on providing tools and information to advertisers, as well as restricting advertising availability on accounts and websites that propagate disinformation. The ACMA is concerned that these provisions focus exclusively on providing such tools and not on the responsibility of platforms to scrutinise advertisements placed via their systems.

These provisions are weaker than those in the EU code, which state that signatories will use commercially reasonable efforts not to accept remuneration from, or otherwise promote accounts and websites that consistently misrepresent information about themselves. The EC's guidance on reforming the EU code seeks to strengthen these provisions. In particular, it states that platforms should commit to 'tighten eligibility requirements and content review processes for content monetisation and ad revenue share programmes on their services'.⁹⁹

The code could also establish cooperative structures for exchanging information on sources of monetised disinformation. This could include a common repository of rejected advertisements, as noted in the EC's *Guidance on Strengthening the EU Code of Practice*.

⁹⁸ European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021, p. 14.

⁹⁹ European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021, p. 7.

Empowering users to identify the quality of news and information

The code includes a range of both proactive and enabling measures to allow users to make more informed choices of news and factual content.

Importantly, it includes broader measures that go beyond the immediate digital platform environment, such as the promotion of digital literacy and the provision of financial support or sustainable partnerships with fact-checking organisations. Several stakeholders have noted the importance of developing sustainable partnerships.

More-detailed commitments to improve the online information environment could include measures that address the role that system design can play in the propagation of disinformation and misinformation, and a commitment to increased transparency over the criteria that platforms (or their algorithms) use to assess the quality of information and to prioritise content.

Many stakeholders have expressed the need for the code to address algorithmic transparency, and this was a concern raised in the DPI.¹⁰⁰ The EC guidance on reform of the EU code also calls for increased transparency in the criteria used for prioritising or de-prioritising information, with the option for users to customise ranking algorithms.

Other, more-structured measures could include frameworks to establish collaboration with experts in media literacy and safety by design, and with fact-checking and news-industry bodies. These would provide channels for expert advice and include agreements to collaborate on the development and implementation of tools and initiatives to improve users' ability to navigate the online information environment. Several stakeholders have noted that there is inadequate recognition in the code of the importance of working closely with news industry bodies or news media.

Collaboration with researchers, government and other stakeholders

The code includes a guiding principle recognising the importance of industry support for independent research.¹⁰¹ Provisions for strengthening the public understanding of disinformation and misinformation through the support of strategic research are also included under Objective 6. Stakeholders have also observed that a similar outcome is included in the EU Code but that progress on this outcome has been slow.

As noted by several stakeholders, more detailed and structured commitments would be valuable in promoting improvement and transparency in this area. These could include frameworks to establish ongoing collaboration with researchers, government and other stakeholders, and could address such matters as data-sharing arrangements to facilitate research, real-time monitoring of disinformation and misinformation or the formation of expert advisory bodies on matters of concern to the public, government, or other industries.

¹⁰⁰ See submissions on the draft code by Australian Strategic Policy Institute, Centre for Responsible Technology, Reset Australia, Digital Rights Watch, the Australian Muslim Action Network, RMIT ABC Fact Check and SBS.

¹⁰¹ Code provision 2.6.

Finding 29: The output-focused framing of several objectives and measures may encourage reporting to focus on outputs rather than progress towards code outcomes.

Finding 30: The code should include industry-wide frameworks for the development and implementation of individual platform measures. Examples could include frameworks to establish:

- > criteria for the assessment of harm
- > criteria for assessing news and information quality
- > processes for the exchange of information between platforms on disinformation and misinformation risks
- > commitments to address the propagation of disinformation and misinformation via platform advertising channels
- > commitments to address the risks of propagation via platform algorithms and architecture.

4.4. Code administration and reporting regime

DIGI will be the administrator of the code and will establish a sub-committee to meet at 6-monthly intervals to review the actions of the signatories and monitor how they are meeting their commitments under the code.¹⁰² The sub-committee will include independent members as well as signatory representatives.

This is an important measure and a sign that the industry is prepared to move towards greater public transparency and accountability. However, no detail is provided in the code on how the sub-committee will operate, including how the independent members will be selected, or the circumstances which might lead a signatory's actions to be considered by the sub-committee.

There is also very little detail on enforcement mechanisms. Provision 7.4 states that signatories agree to develop and document a process describing circumstances in which a non-compliant signatory may be removed. This is an important consideration but removing a signatory may impact the effectiveness of the code as a self-regulatory mechanism. A range of enforcement mechanisms beyond removal should be considered.

The lack of detail on code administration limits the ACMA's ability to assess the likely effectiveness of the code. The ACMA considers that the inclusion in the code itself of a framework setting out principles for the structure and operation of the sub-committee would provide greater transparency and accountability.

Code reviews

An initial code review will take place after 12 months of operation.¹⁰³ Subsequent reviews will occur at 2-yearly intervals. These reviews will be based on the input of the signatories, relevant government bodies (including the ACMA) and other interested stakeholders, including academics and representatives from civil society active in this field.

The ACMA considers the reviews are appropriately timed. The initial review will allow feedback to be considered on the content and structure of the code, the adequacy of platform reporting, the extent of signatories' commitments to the code and the

¹⁰² Code provision 7.5.

¹⁰³ Code provision 7.6.

development of code administration processes. This includes the opportunity to incorporate findings from this report and for other developments to be considered.

The reviews would benefit from the input of a broad range of stakeholders. They should include a public consultation process that is promoted through appropriate channels including on signatories' digital platform services. They should also take into account the input of stakeholders from related industries, including news media and health.

Code complaints mechanism

The code includes a commitment to establish a facility for addressing code non-compliance, including a mechanism for handling unresolved complaints.¹⁰⁴ The complaints facility will be established within 6 months of the commencement of the code. The facility will hear appeals of complaints of code breaches that have not been acted upon by signatories, but not individual complaints of signatories' decisions regarding content on their platforms, including whether specific items of content should be retained or removed.

The ACMA considers this to be a suitable approach for addressing general complaints about code compliance. However, the ACMA is concerned that the code does not place obligations on signatories to have robust internal complaints processes to address user complaints. The ACMA expressed this expectation in its position paper.¹⁰⁵

The code should set an expectation that signatories will have an internal complaints process for matters covered by the code that is transparent, responsive and accessible. Complainants should access this internal process in the first instance. If they are unable to resolve the complaint internally, signatories should provide access to external dispute resolution, so that the matter can be considered and resolved by an independent third party at no cost to the complainant.

The government is committed to considering an external dispute resolution scheme as part of its response to the DPI (outlined in Appendix E). Platforms may wish to consider whether this is an appropriate forum to address escalated complaints under the code.

The code should also set out how complaints, including escalated individual complaints, may be referred to the complaints facility established under the code. It is important to note that there is a point at which individual signatory decisions about content may become a matter of code compliance. For example, where a platform's decision on a piece of content, or the process it follows in making that decision, do not comply with the platform's published policies, or where the policies themselves do not provide adequate protections in accordance with the code.

Performance reporting framework

Signatories are required to submit annual reports that set out progress towards achieving code outcomes and will be published on the DIGI website.¹⁰⁶

Appendix 2 to the code provides a template for an initial report that signatories are required to submit within 3 months of signing up to the code. This template provides a workable foundation for platform reporting. Key areas where the reporting template could be improved include:

¹⁰⁴ Code provision 7.4.

¹⁰⁵ ACMA, [Misinformation and News Quality on Digital Platforms in Australia](#), p. 25.

¹⁰⁶ Code provision 7.3.

- > a clear format to set up existing measures, proposed measures, and performance reporting separately under each outcome
- > a clear requirement for performance reporting to provide adequate data to measure platform performance against each outcome and not just describe the actions platforms have taken. For example, outcome 1(c) in Appendix 2 currently only asks that signatories include links to published policies, procedures and guidelines
- > a more detailed discussion about future plans against each outcome in the report, which would provide greater visibility about changes over time
- > a clearer distinction between the identification of relevant measures or actions (that is, what steps are individual signatories committing to do under the code) and ongoing reporting on the effectiveness of these measures in addressing the code outcomes (for example, how individual signatories intend to demonstrate their measures or actions have been successful).

A more detailed discussion on the application of the template is contained in Chapter 5.

Within 6 months of code commencement, signatories will develop and implement an agreed format for annual reports and a guideline that will inform the data and other information to be included in those reports.

The development of a robust reporting framework is critical to the effectiveness of the code. This development would benefit greatly from broad input and collaboration, including with academic experts and stakeholders in related industries. The guideline should also include the development of standard key performance indicators against each outcome. These are crucial to improving the transparency of signatory actions, encouraging industry progress towards code objectives and monitoring the code over time.

While the ACMA acknowledges the commitment to develop a guideline, without one it is difficult for us to provide a full assessment of the code's reporting framework. The ACMA would welcome the opportunity to provide input to signatories on their proposed guidelines and reporting framework, including collaborating on identifying key metrics and indicators. Some initial guidance is provided at Appendix F.

Finding 31: The code provides a high-level code administration framework. Given that detailed arrangements for code administration, compliance with the code, and consumer complaints are still under development, the ACMA's ability to assess their practical effectiveness is constrained.

Finding 32: The code should include a framework setting out principles for the structure and operation of the sub-committee to provide greater transparency and accountability.

Finding 33: The reporting template provides a workable foundation for the reporting guideline. Reporting should incorporate adequate data to measures performance against KPIs under each outcome; detailed action plans to address areas identified for improvement; and a clearer distinction between measures (that is, outputs) and the effectiveness of these measures (progress towards outcomes).

Finding 34: The proposed 12-month review will provide an opportunity for findings from this report, and other developments, to be incorporated into the code.

Finding 35: The lack of detail on code administration matters, including on the operation of the sub-committee and guidelines for future code reporting, has limited the ACMA's ability to undertake a full assessment on the likely effectiveness of the code.

5. Assessment of platform performance

As a mandatory commitment of the code, each signatory is required to provide an annual report to DIGI setting out its progress towards achieving the code objectives it has opted-in to.

While the code has not been operational for long, the ACMA has met with all code signatories, reviewed their initial annual reports, and is able to draw some preliminary observations on performance based on this information.

This chapter provides a thematic analysis of signatories' initial annual reports, covering:

- > how signatories have met their reporting obligations under the code
- > the range of measures signatories have identified to meet their commitments
- > the effectiveness of these measures (noting the short time the code has been in place and limited amount of data provided by signatories)
- > general reporting requirements and code commitments.

Appendix B includes individual assessments of signatories' reports.

5.1. Platform commitments

The first stage of the ACMA's assessment was to examine platform commitments and whether signatories had met their general reporting requirements under the code. The code prescribes a reporting template, asking signatories to provide information on their business, measures against the relevant outcomes the platform has opted-in to, approach to monitoring performance, and information about future trends.

Opt-in nomination forms and initial annual reports by signatories were published on DIGI's website on 22 May 2021, 3 months after the commencement of the code.

An outline of signatories' opt-in commitments is provided in Table 8 below.

It should be noted that the table is included for clarity only. On the whole, the ACMA considers that where signatories have chosen to opt out of particular outcomes they have done so on justifiable grounds. Some comments on particular decisions are included in individual platform assessments in Appendix B.

Table 8: List of signatories' code commitments

Outcome	Adobe	Apple	Facebook	Google	Microsoft	Redbubble	TikTok	Twitter
(Mandatory) 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.	✓	✓	✓	✓	✓	✓	✓	✓
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.			✓	✓	✓	✓	✓	✓
1c: Users can report content and behaviours to signatories that violates their policies under 5.10 through publicly available and accessible reporting tools.			✓	✓	✓	✓	✓	✓
1d: Users will be able to access general information about signatories' actions in response to reports made under 5.11.		✓	✓	✓	✓		✓	✓
2: Advertising and/or monetisation incentives for disinformation are reduced.		✓	✓	✓	✓	✓ ¹⁰⁷	✓	✓
3: The risk that Inauthentic User Behaviours undermine the integrity and security of services and products is reduced.			✓	✓	✓	✓	✓	✓
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.	✓	✓	✓	✓	✓		✓	✓
5: Users are better informed about the source of political advertising.			✓	✓	✓		✓	
6: Signatories support the efforts of independent researchers to improve public understanding of disinformation and misinformation.		✓	✓	✓	✓ ¹⁰⁸	✓	✓	✓
(Mandatory) 7: The public can access information about the measures Signatories have taken to combat disinformation and misinformation.	✓	✓ ¹⁰⁹	✓	✓	✓	✓	✓	✓

¹⁰⁷ Redbubble has opted in to implement policies and processes that aim to disrupt advertising and/or monetisation incentives for disinformation (5.14) but has opted out of committing to examples under 5.14 as no third-party advertising is permitted to be published on Redbubble. Redbubble does not sell media space to any third-party businesses (5.15, 5.16).

¹⁰⁸ Microsoft has not yet committed to opt in to convening an annual event to foster discussions regarding disinformation within academia and civil society (5.27).

¹⁰⁹ Apple will make and publish an annual report, as a requirement under the code (5.28), but initially will not publish additional information detailing their progress or additional commitments they have made under the code, such as additional reports or public updates (5.29, 5.30).

ACMA commentary

All of the signatories met the stipulated timeframe outlined in the code, providing DIGI with their opt-in nomination forms and interim annual reports within the 3-month deadline. While Adobe and Apple were not initial signatories, it is pleasing to see that they were able to meet this timeframe.

Most of these reports contain information against each of the categories stipulated in the template. However, signatories have taken varied approaches to structuring and formatting their reports. Future reporting would benefit from an agreed, uniform approach across all signatories, to aid direct comparison.

A key focus for the ACMA was on whether signatories had fully articulated their rationale for not opting into specific code outcomes. For the most part, we considered these explanations to be relatively clear; however, in some cases they were not sufficiently targeted to the outcome. For example, Apple did not state why it chose not to opt-in to providing additional information under Objective 7, and Adobe provided a general justification that did not explain why it chose not to opt-in to seemingly relevant outcomes, like supporting strategic research under Objective 6. By contrast, Redbubble offered clear justifications for its decisions, including that it does not disseminate news or permit advertising, and that it is a small company with limited resources for reporting additional information outside of its annual report.

Most signatories were also clear about which of their products and services would be covered by the code. However, there were instances where reports refer to secondary products and services, such as ad service technologies and enterprise software, and it was not always clear why they were included or whether signatories intended for them to be covered. For example, Microsoft reported that all its consumer-facing services will be covered without providing a specific list of services, Facebook included information about measures on excluded services, and Adobe's commitments related to the use of a technology rather than a specific product or service.

The ACMA considers that signatories should be required, in future annual reporting, to provide a list of which services and products are covered by the code. Signatories should also justify the exclusion of any major product or service offering on the basis of specified criteria. Where appropriate, this justification should include the provision of data demonstrating that there is no serious risk of harm associated with the propagation of disinformation or misinformation on that product or service.

Finding 36: On the whole, signatories have met the initial reporting requirements set out in the code.

Finding 37: For the most part, signatories have provided appropriate explanations where they have not opted-in to specific commitments.

Finding 38: A more uniform approach to reporting would assist in cross-platform assessment and increase transparency of platform measures and performance.

Finding 39: For future reports, signatories should clearly specify the products and services covered by the code, and justify any major exclusions.

5.2. Suitability of platform measures

The next stage of the ACMA's assessment was to evaluate the appropriateness or suitability of the measures identified by signatories. This included consideration of whether the measures are appropriately targeted to their relevant code outcome(s), including a clear nexus to Australia or Australian users, and a commitment to continual improvement and future initiatives. We also examined whether key concepts in the code had been applied in a consistent manner across signatories.

In assessing the suitability of these measures, we must first recognise the dynamic nature of disinformation and misinformation. Over the last 18 months, the number of platform measures has expanded across industry, largely in response to the COVID-19 pandemic. This demonstrates a responsiveness to changes in the information environment, but also raises questions as to whether these efforts are temporary or signal stronger industry-wide action to address the challenges of disinformation and misinformation.

While this discussion is primarily based on the information contained within signatory reports, the ACMA has also separately been monitoring changes in platform policies over the last 18 months, as reflected in the timeline at Appendix C.

ACMA commentary

Interim annual reports show that signatories have adopted a wide range of measures to address the problems of disinformation and misinformation and to improve the quality of news and information on their services.

Most of the measures identified in the initial reports are pre-existing and based on global, platform-wide policies and community standards, rather than new measures driven by the introduction of the code or that expressly target Australian users. This is not unexpected, particularly as the code is a new initiative, and Australia is a relatively small market.

Notwithstanding this, some signatories did identify measures they have implemented, or are planning to implement, that focus on Australia:

- > signatories including Facebook, Google and TikTok discussed their partnerships with local fact-checking organisations
- > several signatories, including Apple, Facebook, Google, Twitter and TikTok, provided information on measures they had implemented in response to events directly relevant to Australia, including dedicated spaces to prioritise quality news and locally sourced authoritative information relating to high-risk matters including COVID-19 and natural disasters.

Future initiatives

One disappointing aspect of the reports was the limited discussion about future measures or initiatives that signatories are planning to introduce under the code. In general, the reports heavily focused on current measures and past actions, and signatories mostly avoided providing any concrete information on planned developments or future initiatives. The one exception was Facebook, which included information on its plans to expand its policies to improve transparency of the source of political advertising to issues-based advertising in Australia.

While we recognise the challenges in providing industry and bad actors with advance notice of changes, it is important for signatories to signal any broad commitments that are planned to address misinformation. This would strengthen future reporting and better enable stakeholders to track the rollout of new measures under the code.

Interpretation and use of code terms

The reports show inconsistencies in the interpretation and use of key code terms between signatories. In discussing measures, several signatories referenced definitions of 'disinformation' and 'misinformation' from their internal policies and community standards, rather than the agreed definitions under the code. This presents a fundamental issue in the construction of the code, hampering assessment of signatories' performance against the code and any industry-wide comparisons.

Similar issues have been raised within the context of the EU code, with recent European Commission guidance recommending a harmonised template that allows, to the extent practicable, cross-platform comparisons.

Some examples of this include:

- > Facebook used different definitions of disinformation and misinformation from the code, with disinformation being used to refer to inauthentic behaviour with the intention to deceive, and misinformation to refer to content that is false or misleading. This reflects Facebook's existing policies and the different tools it uses to address these 2 problems. Despite this, Facebook's measures appear to cover the scope of the code.
- > TikTok used a narrower definition of misinformation than in the code, defining it as false or inaccurate content. Content that is misleading is considered only in the context of elections and civic processes, and manipulated media. Disinformation was included as coordinated inauthentic behaviour to exert influence and sway public opinion while misleading individuals and our community about the account's identity, location, or purpose, and was included under misinformation.
- > Google noted that, in practice, it does not make a distinction between disinformation and misinformation in the application of its policies.

The reports also included detail of content and services that are excluded from the scope of the code. For example, Facebook provided some information on its private messaging service, Facebook Messenger, despite these types of services being excluded from the code. The additional information is welcome to provide visibility about signatory actions about misinformation. However, a revised reporting template should make it clear that the provision of this information relates to excluded content and services.

Assessment of harm and proportionality

The way in which signatories assess harm and apply proportionality and risk considerations under their policies is not always transparent in their reports. This could be clearer in future reports. For example, TikTok's misinformation policy does not limit harms to those who pose an imminent and serious threat. However, its approach to assessing the extent of harm, and therefore whether content should be removed or labelled, is not clearly articulated.

Other signatories also reported their approach to harms below the serious and imminent threshold set out in the code. The ACMA welcomes this additional reporting given we consider this threshold to be too high to capture the full scope of potential chronic and acute harms (see Chapter 4). Both Google and Facebook set out graduated approaches to disinformation and misinformation based on the risk of harm. Facebook referred to its COVID-19 misinformation policies, which are transparent about particular topics and claims that it considers to be harmful. It noted that these are established in collaboration with experts in health communication and other fields. Facebook also referred to its policy of reducing content that does not contravene its community standards but has been rated as false by fact-checkers.

User reporting of disinformation and misinformation to platforms

The code sets the expectation that signatories will have functions for users to report all disinformation and misinformation content.¹¹⁰ All signatories who opted into this provision provided descriptions of how users can report content that contravenes their policies. Google, Microsoft and Redbubble also provided screenshots of their existing reporting functions. In all cases, signatories provide tools to allow users to report content via their platforms.

However, not all signatories allow users to report content under all of their disinformation and misinformation policies or for all relevant aspects of their services. For example, Twitter users are only able to report content under its platform manipulation and spam policy, but not under its COVID-19 misleading information policy. Facebook allows users to report content on Facebook and Instagram as ‘false information’, which will be assessed under its misinformation policies. It does not consider that users would be able to detect inauthentic behaviour, and as such does not provide reporting mechanisms for suspected disinformation.

The code also sets the expectation that signatories will publish general transparency data on the actions taken in response to user reports and the reporting template asks platforms to provide information on how they do this.¹¹¹

In most cases, signatories reported that they provide general information about their responses to user complaints in their public transparency reports. Transparency reports provide aggregated data on signatories’ enforcement of their policies and, in some cases, provide information on user complaints data. For example, YouTube reports on videos removed by source of first detection – automated flagging, user, trusted flagger, NGO or government agency – as well as on videos removed by removal reason – harmful or dangerous, harassment, hateful or abusive, violence and violent extremism and other. However, information of this kind as it relates specifically to the code was not included in any signatories’ reports.

On the whole, signatories provided limited information on their procedures for processing reports from users, including whether users could expect to receive a response. Twitter reported that users would receive a response directly from their support teams about the results of any investigation or enforcement action. TikTok reported that creators are notified when their content has been found to violate its policies but did not comment on what information is received by users who make a complaint. Although not included in their initial report, YouTube users are able to view their reporting history and the action that was taken by YouTube in response to the report. Redbubble told the ACMA that it provided an automated message in response to user reports, but that it would not contact the user directly unless it needed more information, or if the user had reported the content via social media.

From discussions with signatories, it is clear that in many circumstances, the onus will fall on the user to check whether the platform has taken action against reported content, rather than receiving direct notification. This lack of information is consistent with the DPI finding that there is substantial room for improvement in digital platforms’ user reporting and internal dispute resolution processes, including by increasing transparency and consumer access.¹¹² Signatories should look to increase transparency by improving the information they provide to users about the outcomes of their reports.

¹¹⁰ Code provision 5.11.

¹¹¹ Code provision 5.13.

¹¹² ACCC, [Digital Platforms Inquiry Final Report](#), 2019, pp. 507–9.

Need for further clarity and specificity

On the whole, signatories provided concrete and detailed descriptions of their measures under each outcome. In some cases, however, it was not completely clear how certain measures will contribute to the achievement of the code outcomes under which they have been reported.

For example, TikTok reported its Asia–Pacific Safety Advisory Council as a measure against outcome 6, which aims to improve public understanding of disinformation and misinformation. It is not clear, however, whether the council's findings or advice will be made publicly available, thereby improving public understanding of disinformation and misinformation. Outside its report, TikTok has advised that the committee will provide advice to TikTok on its content moderation policies, which suggests it is a measure targeted at Objective 1.

The case study below provides a high-level assessment of the reported measures against Objective 4. Signatories reported a range of measures that appear to be relevant to achieving the objective. However, signatories' reporting was not detailed, in terms of either quantitative or qualitative data, making it difficult to assess the relevance and impact of measures on users. This is a common theme across the reports provided.

Case study: Assessment of reported measures against Objective 4

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

Signatories reported a broad range of measures intended to enable users to make more-informed choices about the news and factual content they access via digital platforms and to identify misinformation.

Facebook, Twitter, TikTok and Google reported on their financial support for media literacy programs.

Google, Microsoft and Facebook outlined measures to prioritise authoritative information in search results. For example, Google and Microsoft reported making ongoing improvements to promote authoritative sources and demoting borderline or low-authority sources in Google Search and Bing. Microsoft also reported on a specific feature, Intelligent Search, which is designed to promote a diversity of perspectives by displaying 'all valid answers to a question' in a carousel.

Apple, Microsoft and Google reported on their news aggregation services, which provide news content subject to an independent editorial code and complaints scheme. Google's Full Coverage feature in Google News also provides videos and articles from different publishers on a news story selected by the user. Apple's and Microsoft's services both involve human editorial input.

Many signatories reported employing technologies to signal the credibility of news sources. Facebook, Google and Twitter label government-controlled channels and accounts, and Apple maintains brand and logo information on individual articles in Apple News. Signatories also outlined measures to assist users or platforms to check the accuracy of online news content or identify its provenance. For example:

- > Apple News and Bing News both partner with NewsGuard, a news-rating organisation
- > Facebook, Google, Microsoft, Twitter and TikTok all partner with fact-checkers and apply labels to fact-checked information
- > Adobe is developing a system for creators and publishers to embed visual and audio-visual content with attribution data as part of its CAI.

A number of measures reported by signatories against Objective 1 are likely to contribute to helping users identify misinformation and therefore are relevant to Objective 4. These include prompts to authoritative sources when searching for significant social topics, such as COVID-19. Facebook also reported a feature to direct users to authoritative sources if they have previously encountered COVID-19-related content subsequently identified as false by fact-checkers.

Across their reports, signatories have reported a range of measures that appear to be relevant to achieving Objective 4. However, signatories' reporting is not detailed, in terms of either quantitative or qualitative data, making it difficult to assess the relevance and impact of measures on users.

Broadly, we observe that signatories could strengthen commitments by implementing further measures in collaboration with news providers. For example, credibility signalling and fact-checking initiatives would be enhanced by a more holistic approach. This was also raised by stakeholders who considered that news quality was a primary mechanism for countering misinformation. Stakeholders also suggested that credibility signalling of news sources could be enhanced by displaying logos of relevant journalistic or industry bodies.

- Finding 40:** Signatories have a wide range of measures in place to address the problems of disinformation and misinformation and to improve the quality of news and information on their services. They also demonstrate responsiveness to significant changes over the last 18 months, as well as to public and government calls for stronger action.
- Finding 41:** It is expected that signatories will develop more Australia-focused measures over time.
- Finding 42:** In general, the reports are heavily focused on current measures and past actions, and signatories have provided little systematic information on future initiatives. In some cases, it is not clear to what extent certain measures will contribute to the achievement of the code outcomes under which they have been reported.
- Finding 43:** There are inconsistencies in the interpretations of key terms between signatories, which are drawn from pre-existing definitions from their internal, often global, policies. This makes it difficult to interpret and assess performance and to make industry-wide comparisons.
- Finding 44:** A harmonised template would assist in comparing initiatives across platforms. It would also allow clear reporting on additional information beyond the requirements of the code.
- Finding 45:** Signatories have provided a large range of information on the actions they have taken to address misinformation, disinformation and news quality and to invest in collaborative initiatives. This demonstrates signatories' commitment to addressing these issues.
- Finding 46:** The information signatories have provided is heavily focused on platform outputs and on volumetric data. Reporting lacks systematic data, metrics or key performance indicators (KPIs) that establish a baseline and enable the tracking of platform and industry performance against code outcomes over time.

5.3. Effectiveness of platform measures

An important component of an outcomes-based approach is the ability of entities to demonstrate the effectiveness of the measures they have adopted to meet the identified outcomes. It follows that entities must be able to provide sufficient information and data about their respective measures in order to baseline, monitor and track their performance against the stated outcomes.

While the code has not been in operation for long, the ACMA has considered some preliminary findings on the effectiveness of signatories' measures against their respective objectives.

Measures to address misinformation and news quality

Reports provided a large range of information and data on the actions that signatories have taken to address misinformation, disinformation, and news quality. This gives an indication of the nature and extent of these issues on their services. It also demonstrates the signatories' commitment to addressing these issues via a variety of measures and initiatives:

- > Google, Facebook, Twitter and TikTok provided information on their transparency centres and transparency reports, which collate information and data on the actions they are taking to address online safety and authenticity.
- > Google and Facebook provided recent global data on the number of accounts removed for engaging in coordinated inauthentic behaviour and influence operations.
- > Facebook, Google and TikTok provided data on the number of takedowns of potentially harmful misinformation relating to COVID-19, including Australia-specific data.
- > Google provided a case study on its approach to COVID-19 misinformation, including raw data on impressions of authoritative information panels.

Several signatories provided data on their investment in measures and initiatives to combat disinformation and misinformation and improve the quality of online news and information:

- > Google provided expenditure data relating to the Ad Grants Crisis Relief program to help government agencies and global NGOs run critical public service health announcements during COVID-19. It also provided funding data on collaborative initiatives such as the Alannah and Madeline Foundation's Media Literacy Lab and the number of students currently enrolled.
- > Facebook provided expenditure data on a global funding round for academic research into misinformation and polarisation. Two winners were located at Australian universities. Facebook also provided expenditure data on grants provided to their fact-checking partners to improve capacity during COVID-19.

Need for further data and metrics

The information signatories have provided is heavily focused on platform outputs – measures that signatories have put in place – and on volumetric data relating to the implementation of those measures. Where signatories provided comparative or success-oriented data, this was often piecemeal or not directly related to actions under the code:

- > Google noted that changes to YouTube recommendation systems resulted in a 70% reduction in time watching non-subscribed, recommended content in the US in 2019. They also noted that their aim is to have views of non-subscribed, recommended borderline content (that is, content close to breaching content policies) below 0.5%. However, no relation was drawn between these figures.
- > Google provided data on YouTube video takedowns that showed the percentage of videos removed before they were viewed and, separately, the percentage removed for spam, scams or misleading content. While this may assist in establishing a baseline, it does not cross-reference these 2 data points.
- > Twitter's report provided raw data on account actions, suspensions and deletions relating to its misleading information policy. These will assist in establishing a baseline for future reports but do not establish comparative metrics.

On the whole, reporting lacked systematic data, metrics or key performance indicators (KPIs) that establish a baseline and enable the tracking of platform and industry performance against code outcomes over time. Similar criticisms have been made about the EU Code, and the European Commission has recommended the development of both platform-specific and industry-wide KPIs. The ACMA has also developed some high-level guidance at Appendix F to assist DIGI and signatories ahead of their next annual report.

Australia-specific data

Signatories provided some data on the Australian context, but this was again piecemeal, or not directly related to actions under the code.

- > Twitter provided information on account suspensions and deletions for Australian accounts but only global data on content labelling. Twitter noted that it has not generally identified specific, large-scale, targeted information operations originating outside of Australia and targeting people and conversations within Australia.
- > TikTok provided Australia-specific data on content labelling and removals and views of the COVID-19 information hub (relevant to outcome 1a) but did not provide data against other outcomes.
- > Facebook provided data on the number of pieces of content removed from Australian-specific pages and accounts between March and December 2020.
- > Google provided data on the number of ads they blocked from Australian-based advertisers for violating their misrepresentation ads policy. This includes ads violating misleading representation, clickbait, and unacceptable business practices policies. Google also provided raw figures on ads blocked or removed from Australia-based advertisers relating to COVID-19. This included ads removed for misleading claims, but also non-code-related matters such as price-gouging.
- > Google provided data on the number of accounts removed for engaging in coordinated influence operations. These are global removals and no information was provided on the extent to which such operations affect Australian users.

While disinformation and misinformation are global issues, and signatories operate on an international scale, reporting should include Australia-specific data and signatories should establish a reporting regime against the Australian code.

The ACMA acknowledges that some signatories have provided confidential data points to inform this report. While this has assisted in the development of this report, we consider this data should have been included in published reports.

Trends in data

The reporting template asks signatories to provide qualitative or quantitative data on trends where available. As noted above, there were some examples of case studies where signatories provided empirical data as backing for the success of their measures. For example, Redbubble reported data on the numbers of sales made for content tagged by uploaders with anti-vaccination tags since April 2015.

However, there was generally a lack of long-term data in signatory reports, particularly longitudinal quantitative data about the prevalence and types of disinformation and misinformation or other key metrics, such as user behaviour.

Trend-related data would contribute to a greater understanding of the extent and impact of disinformation and misinformation in Australia. Importantly, systematic treatment of such data would create greater transparency in the effectiveness of signatories' actions in addressing emerging issues.

Finding 47: Reports provide some data on the Australian context, but this is often piecemeal or not directly related to actions under the code. Reporting should include Australia-specific data and signatories should establish a reporting regime against the Australian code.

Finding 48: Reporting lacks trend-related data. Trend-related data would contribute to a greater understanding of the extent and impact of disinformation and misinformation in Australia.

6. Considerations for future reform

Over the last 18 months, signatories have demonstrated their commitment to addressing disinformation and misinformation by implementing a wide range of measures and collaborating on the development and introduction of the code. However, while the code enjoys broad support across industry, it is too early to draw any concrete conclusions on its overall impact or effectiveness.

The building blocks for an effective self-regulatory scheme – such as complaints-handling processes and other code administration matters – are still under development. Broader public confidence is unlikely until the industry can demonstrate it has built the necessary frameworks to effectively self-regulate.

As the code is still in its infancy, and due to be reviewed by signatories in early 2022, the government could wait and see how the industry responds over the next 12 months. However, it may also wish to consider taking steps to strengthen regulatory oversight and proactively address some of the issues identified in this report to increase the overall likelihood of the code's success.

In this chapter, we put forward recommendations to assist government in considering its approach.

6.1. Consideration of findings

This report has made a range of findings on the code development process, the code content and code framework. These include:

- > The need for **greater publicity** when consulting on changes to the code, including promotion through public communications channels and engagement with the media.
- > The suggested move to an **opt-out code framework** for code outcomes. Signatories could provide statements against outcomes that do not apply to their business model or services, rather than opting-in to each outcome individually.
- > **The scope of the code is limited** by the narrow definition of harm and the exclusion of some relevant products and services. This presents a risk that the code will fail to provide adequate safeguards against the full range of potential harms caused by the propagation of disinformation and misinformation on digital platforms.
- > **The code reporting and administration framework is high level and yet to be developed.** Additional guidance outside of the code is yet to be developed, which will be vital to the effectiveness of the code.

DIGI and signatories are encouraged to consider these findings when finalising their code administration framework and reviewing the code in February 2022.

Recommendation 1: The government should encourage DIGI to consider the findings in this report when reviewing the code in February 2022.

6.2. Continued oversight

The ACMA's position paper provided a useful framework to guide the code development process and the draft code. In releasing the draft code for consultation, DIGI stated it had regard to this paper in the development of the code. Several submitters also drew upon the paper in their submissions to the draft code – suggesting elements of the ACMA's guidance (such as expanding the code to cover misinformation) should be included in the final draft. Subsequent discussions with submitters indicated that the paper provided a good framework for distilling the issues in an Australian context and was useful to inform the development of their submissions.

As discussed in this report, the delay in finalising the code had follow-on effects on the finalisation of the code administration framework and reporting guideline. Both the framework and the guideline are still under development and are scheduled to be in place by the end of August 2021. Recent discussions with DIGI indicate that governance design is focused on incentivising signatories to meet the commitments under the code and drive improvements over time. An independent committee with defined terms of reference will be established to make sure signatories are meeting their code commitments and will also have a role in the independent code review. DIGI has advised that a key component of the model will be on arrangements to support transparency reporting.

While DIGI has kept the ACMA updated on its progress, the information available is extremely preliminary. This reflects the need to develop a completely new function by an industry with limited experience in codes. More time and concrete information is required for a considered assessment of code administration arrangements.

Given the status of code oversight arrangements, there is an argument that continued oversight of the code is warranted. A continued oversight role would also allow the ACMA to provide a more robust and fully formed assessment of the effectiveness of the code once it has been in operation for a reasonable period of time. This is also consistent with the government's response to the DPI, which suggested the ACMA's June 2021 report would be the first of a number.

Recommendation 2: The ACMA will continue to oversee the operation of the code and should report to government on its effectiveness no later than the end of the 2022-23 financial year. The ACMA should also continue to undertake relevant research to inform government on the state of disinformation and misinformation in Australia.

It is possible that the ACMA's findings will be addressed in the February 2022 code review. However, many of the code deficiencies identified in this report were raised with DIGI and the signatories throughout the code development process. DIGI's experiences in developing the code also suggest that it may be difficult to achieve consensus on a range of code issues amongst current signatories.

The ACMA considers there is a continued role for guidance to publicly articulate clear expectations where appropriate. Given the influential role that the ACMA's position paper had in the development of the draft code, this role should be continued.

The issuing of guidance has also been used in other jurisdictions to influence the development of voluntary codes. For example, the European Commission recently

issued *Guidance on Strengthening the Code of Practice on Disinformation* to articulate its expectations to platforms on the proposed revision of the EU Code.¹¹³



Strengthening the EU Code

The European Commission issued its guidance in May 2021, following its formal assessment of the EU Code in September 2020.¹¹⁴ The guidance sets out how platforms should step up their measures to address the shortcomings identified in the assessment of the EU Code.

Some of the key recommendations from the guidance include:

- > an expansion of the code to include misinformation in some areas, as well as private messaging services
- > wider participation from both established and emerging platforms, as well as fact-checkers, content assessment organisations, and technology developers
- > wider participation from both established and emerging platforms, as well as fact-checkers, content assessment organisations, and technology developers
- > a requirement for signatories to publicly justify their reason for opting out of certain code provisions
- > enhanced code provisions on the demonetisation of disinformation on advertising channels.

It also suggests an enhanced monitoring and reporting framework which includes the development of:

- > platform-specific and industry-wide KPIs
- > a harmonised reporting template to allow for cross-platform comparisons
- > publicly accessible transparency centres created by platforms
- > a new data access framework for the research community
- > a dedicated taskforce consisting of representatives from signatories, European Digital Media Observatory (EDMO), European Regulators Group for Audiovisual Media Services (ERGA) and other relevant experts, to evolve and adapt the code to technological, market, and legislative developments.

The Commission will continue to oversee the strengthening the EU Code, with a first draft anticipated in the European autumn 2021.

The continued oversight role will also allow the ACMA to assist signatories in the development of more robust reporting arrangements. Improved reporting arrangements also need to be supported by research to monitor the impacts of the problem and better understand the effectiveness of platforms' measures. The ACMA has a strong existing research capability to understand this work, drawing upon our existing work in the media and communications sector. This report could provide a baseline for further research to monitor and understand this dynamic problem. The continuation of targeted and relevant research would also be valuable to inform

¹¹⁴ European Commission, [Commission Staff Working Document: Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement](#), September 2020.

government of the state of misinformation and disinformation in Australia, and inform future digital literacy initiatives.

6.3. Improving signatory reporting

As noted in Chapter 5, the initial reporting by signatories was inconsistent and, in general, lacked the level of detail necessary to benchmark performance or assess the effectiveness of individual platform measures. Without the identification of KPIs supported by robust data, it will remain difficult for signatories to verify their progress towards agreed code outcomes, and for industry to demonstrate the overall effectiveness of the code.

The EC's recently issued *Guidance on Strengthening the Code of Practice on Disinformation* states the need for an enhanced reporting framework that includes many of the same elements we have identified as lacking in the initial platform reports under the code.¹¹⁵ The EC also calls for specific improvements in general platform transparency, including an obligation on platforms to create online transparency centres that provide data against common industry KPIs.

The ACMA acknowledges that signatories' initial reports represent their first attempt at reporting under the code, and that there are several initiatives underway that could improve reporting over time. Signatories are currently working towards agreement on a consistent annual reporting format and approach, and there will be further opportunities to review the strength of reporting processes during the upcoming 12-month code review. The ACMA encourages platforms to consider establishing more-formal collaboration processes to inform these reviews. The ACMA would welcome the opportunity to provide additional guidance on these issues as the reporting guideline is developed. Some initial guidance on reporting and measurement issues has been provided at Appendix F.

However, the experiences of the EU code suggest this is a challenging issue, and that signatories are likely to need further incentives to increase transparency. As multinational corporations, internal policies and priorities may limit the extent to which platforms are willing to invest in the necessary systems and resources to allow for more detailed, Australia-specific reporting. Some signatories have indicated to the ACMA that stronger regulatory backing would give them certainty and justify the allocation of resources to generate reporting capability in new areas.

Recommendation 3: To incentivise greater transparency, the ACMA should be provided with formal information-gathering powers (including powers to make record keeping rules) to oversee digital platforms, including the ability to request Australia-specific data on the effectiveness of measures to address disinformation and misinformation.

Formal information-gathering powers would allow the ACMA to better monitor the progress of the voluntary code and incentivise behavioural change across industry through greater transparency.

In France, the national media regulator Conseil supérieur de l'audiovisuel (CSA), has formal information-gathering powers to oversee digital platforms, including the ability to request specific data points and issue recommendations to platforms to improve reporting. Platforms are also required to submit annual performance reports to CSA

¹¹⁵ European Commission, [Guidance on Strengthening the Code of Practice on Disinformation](#), 26 May 2021.

outlining what measures they have taken to address disinformation in France. This complements the voluntary code at the European level.

The ACMA's consultations with the CSA suggest that this approach has improved transparency of platform activities. Providing the ACMA with similar powers would incentivise signatories to voluntarily make improvements to their annual reporting against the code and allow the ACMA to compel individual digital platforms to provide more detailed or robust data on their performance if their annual reporting is insufficient. This information would also help inform future ACMA reports to government on the ongoing effectiveness of the code and identify systemic industry-wide issues that could inform targeted research.

The power to issue requests for specific information would be consistent with the treatment of other industries that the ACMA regulates across the communications sector, such as the ability to obtain information and documents from telecommunications carriers under section 521 of the *Telecommunications Act 1997*, or to obtain documents through investigations powers under section 173 of the *Broadcasting Services Act 1992*.

Additionally, government could provide the ACMA with the ability to make record-keeping rules for the digital platform industry, such as under section 529 of the Telecommunications Act. Targeted record-keeping rules could require all digital platforms to keep and regularly provide the ACMA with consistent and locally relevant data that could be relied upon to monitor and analyse industry-wide changes and developments over time. If this approach were taken, the ACMA would seek public consultation on any proposed record-keeping rules.

6.4. Code administration

Code signatories and DIGI have demonstrated a great deal of commitment in the development of a single, industry-wide code that has attracted strong industry support. However, with the code administration framework not yet fully developed, areas of concern remain.

Some stakeholders have raised questions about DIGI's independence and lack of transparency over its governance arrangements and funding. We note DIGI does not publish any details about addressing conflicts of interests or the process for membership. Transparency in these areas should be increased.

Major platforms are orders of magnitude larger than other regulated entities in the sector and should therefore be prepared to adequately resource voluntary initiatives.¹¹⁶

The ACMA considers that the industry should be given additional time to bed down its voluntary code. However, the risk remains that the current self-regulatory approach may prove insufficient to incentivise broader behavioural change across industry, as:

- > compliance with the current Code is uncertain at this stage given the data provided by platforms
- > it is not certain that current deficiencies with the code will be addressed by the industry in its 12-month review

¹¹⁶ Four of the 8 signatories (Apple, Microsoft, Alphabet (Google's parent company) and Facebook) are within the top 10 most valuable publicly listed companies in the world by market capitalisation, each valued at over \$1 trillion USD. By comparison, Australia's largest telecommunications provider, Telstra, ranks #610 globally; CompaniesMarketCap, [Largest Companies by Market Cap](#), 29 June 2021.

- > there are a range of non-signatories to the Code (see next section)
- > usage of platforms may expand rapidly and new services introduced without these being brought quickly into the Code's remit.

In response, the government may wish to establish a 'fall-back' regulatory framework to enable future intervention, if required to address non-compliance of new issues as they emerge.

Recommendation 4: The government should provide the ACMA with 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 prove inadequate, or the voluntary industry code fails.

The ACMA currently has no regulatory powers to underpin its oversight role. The establishment of a suite of reserve regulatory powers would allow the ACMA to take further action if required. Actions could range from buttressing the current voluntary Code with a registration process to incentivise industry to develop and enforce compliance with codes, through to standards-making powers if a code fails to address those harms. Such reserve powers would be defined by the government and may be confined to issues of most concern.

Developing this reserve power framework would also improve incentives for industry to improve voluntary arrangements and provide an appropriate backstop for further action if required. Care would be needed to ensure the approach is responsive to the risks of disinformation and misinformation, and the concerns of stakeholders. This can be balanced by a robust public consultation process.

Non-signatories to the code

There is a risk that disinformation and misinformation on non-signatory services may become a larger issue in the future. As evidenced in Chapter 2, some platforms have experienced significant growth in Australia over the past 18 months. With major platforms implementing measures to address misinformation and disinformation on their services, purveyors of misinformation are moving to alternative platforms, like Telegram, that are not code signatories and have less stringent content moderation policies.

As the code is voluntary, there is currently no mechanism to compel platforms to sign up to the code. In the first instance, consideration could be given to tailoring the code to provide greater proportionality in reporting and other administrative matters for smaller platforms. This may assist in encouraging these platforms to sign up voluntarily.

More-formal options could be considered for platforms that do not participate in voluntary arrangements or reject the emerging consensus on the need to address disinformation and misinformation. If these platforms continue to grow their user-base, they may present a higher risk to the Australian community.

This is an issue for both industry and government to consider and continued formal monitoring and reporting is therefore recommended.

To address the risk of emerging platforms, the proposed *EU Digital Services Act* uses a co-regulatory backstop that would enable the EC to enforce the participation of particular platforms in a revised EU Code.

Providing the ACMA with reserve code-registration and standard-making powers such as those outlined in Recommendation 4 would provide the government with the option to act quickly to address potential harms on emerging platforms if required.

6.5. Improved collaboration mechanisms

A consistent theme in discussions about online disinformation and misinformation is the complex, dynamic and multi-sided nature of the problem.

While platforms bear considerable responsibility for the quality of the information environment on their services, a one-sided regulatory response that places sole responsibility for addressing the problem on platforms is unlikely to be effective.

At a minimum, ongoing monitoring of online disinformation and misinformation is required to provide oversight of a rapidly changing environment and a dynamic industry where new and significant risks continue to emerge. Work is underway within the Department of Home Affairs and the Department of Foreign Affairs and Trade to fulfil this monitoring role on behalf of government.

However, there is also increasing recognition that formal frameworks for widespread collaboration – that include, but extend well beyond, digital platforms – are critical to addressing the problem. Collaborative frameworks can encourage greater information-sharing between government, industry, academia and civil society to promote increased understanding of the nature and extent of the problem and enable robust assessment of the effectiveness of measures. They can also provide stronger capability to monitor emerging risks and develop consensus-based approaches to addressing them. Several submitters to the consultation on the draft code called for greater government coordination in this area.

In the EU, the European Digital Media Observatory (EDMO) was established to fulfil both a monitoring and coordination role. By providing support to independent researchers and fact-checkers across 8 national hubs, EDMO will increase the capacity to detect and analyse disinformation campaigns across member states. The EC has called for EU Code signatories to establish formal cooperative mechanisms with EDMO.

Recommendation 5: In addition to existing monitoring capabilities, the government should consider establishing a Misinformation and Disinformation Action Group to support collaboration and information-sharing between digital platforms, government agencies, researchers and NGOs on issues relating to disinformation and misinformation.

Given the range of issues identified above, the government may consider the value of establishing a Misinformation and Disinformation Action Group to support collaboration, cooperation and information-sharing on issues relating to disinformation and misinformation. The Group could include digital platforms, government agencies, interested researchers and NGOs. Key areas of focus for the group could include identifying and monitoring emerging risks and data sharing to improve understanding and inform awareness initiatives.¹¹⁷ In particular, cooperation needs to be focused on harms to users and on improving the overall online information environment. This is consistent with approaches taken internationally and coordination with international bodies like EDMO to develop consistent approaches would also be of great benefit.

6.6. Related areas of concern

There are several related areas of concern that government may wish to take into account when considering any further responses to the code.

Misleading financial advertising

The dynamic nature of misinformation and disinformation and the flexibility of the code contribute to a lack of clarity about what the code covers or should cover. For example, recent media reporting suggested that the code would cover misleading financial advertising in the context of the Mayfair 101 investment scheme.¹¹⁸

Signatory reports from Google and Microsoft indicated their advertising policies would capture this type of advertising. However, these policies did not prevent advertising by the Mayfair 101 investment scheme.

The ACMA notes also that the code is intended to address harms arising in specific areas of platform responsibility. As a voluntary, outcomes-based code, it places no obligations on platforms to take particular actions. In addition, the misleading financial advertising in the Mayfair 101 case may not reach the threshold of having a reasonable likelihood of serious and imminent harm to be considered misinformation under the code.

The code may, therefore, not be the appropriate vehicle for establishing protections in areas that are not directly related to the online environment, where principles such as the freedom of expression are not a primary concern. The government may wish to consider whether this issue should be considered in the context of changes to the *Australian Securities and Investment Commission Act 2001*, Australian Consumer Law or other associated regulation.

Micro-targeted advertising

The 12-month review should consider extending the code to explicitly cover related issues of platform responsibility such as misleading advertising using micro-targeting. A similar extension has also been contemplated in the EC's *Guidance on Strengthening the Code of Practice on Disinformation*.

The ACMA intends to continue monitoring this area to identify and assess any areas of harm that may not be addressed by the code or related initiatives.

¹¹⁷ Montgomery, [Disinformation as a Wicked Problem: Why We Need Co-Regulatory Frameworks](#),

Brookings Institute, 2020, offers a detailed discussion of how such mechanisms might be approached.

¹¹⁸ Davidson, J., '[Review of tech disinformation code promised after Mayfair 101 scandal](#)', *Australian Financial Review*, 12 April 2021.

News quality initiatives

Some news quality initiatives (such as the obligation to develop an original news proposal) are included as part of the minimum standards of the news media bargaining code. These initiatives do not become an obligation on a platform until it is designated under the news media bargaining code. Some stakeholders have indicated concern that news quality-related initiatives may fall between the 2 codes. While there is a range of activities underway targeted at improving news quality (see Appendix E), consideration could be given to addressing any gaps via the provision of formal guidance during the review of the disinformation and misinformation code.

Extension of code beyond digital platforms

Consideration could also be given to extending the scope of the code beyond digital platforms. The EU Code includes advertising bodies as signatories, and the recent EC guidance recommends broader participation from the advertising industry to increase the code's power to drive the de-monetisation of disinformation. It could also be extended to include local fact-checking organisations, content assessment organisations, and those providing tools and solutions for fighting misinformation. We note that Adobe has already signed up to the code, and other technology providers could follow Adobe's lead.

Appendix A: Full list of recommendations and findings

List of recommendations

Recommendation 1: The government should encourage DIGI to consider the findings in this report when reviewing the code in February 2022.

Recommendation 2: The ACMA will continue to oversee the operation of the code and should report to government on its effectiveness no later than the end of the 2022–23 financial year. The ACMA should also continue to undertake relevant research to inform government on the state of disinformation and misinformation in Australia.

Recommendation 3: To incentivise greater transparency, the ACMA should be provided with formal information-gathering powers (including powers to make record keeping rules) to oversee digital platforms, including the ability to request Australia-specific data on the effectiveness of measures to address disinformation and misinformation.

Recommendation 4: The government should provide the ACMA with 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 prove inadequate, or the voluntary industry code fails.

Recommendation 5: In addition to existing monitoring capabilities, the government should consider establishing a Misinformation and Disinformation Action Group to support collaboration and information-sharing between digital platforms, government agencies, researchers and NGOs on issues relating to disinformation and misinformation.

List of findings

Finding 1: Most Australians are concerned about, and have experienced, online misinformation. Higher exposure is associated with heavy use of digital platforms, disproportionately impacting younger Australians.

Finding 2: Access to authoritative and trusted sources of news and information is an important mitigation against misinformation. Those that rely on social media as a main source of news have a greater likelihood of being misinformed about COVID-19.

Finding 3: Given its nature and the ongoing challenges in accessing relevant data, the true scale and volume of misinformation in Australia is currently unknown.

Finding 4: Australians report seeing the most amount of misinformation on large platforms such as Facebook and Twitter. However, private messaging services and smaller platforms with less strict content moderation policies, like Telegram, are also being embraced by conspiracy-oriented communities.

Finding 5: Misinformation typically stems from small online conspiratorial communities, but can be amplified by influential individuals, digital platform design, as well as the media.


- Finding 6:** Conspiratorial content is designed to be highly engaging, fuelling outrage, and building on a sense of community. The confluence of conspiracy theories around COVID-19 has created more paths to online misinformation.
- Finding 7:** There is some evidence of co-ordinated inauthentic activity surrounding popular misinformation narratives in Australia. Those who spread misinformation often seek to reframe global conspiratorial narratives, like QAnon, in a local context.
- Finding 8:** Misinformation narratives can result in a wide range of acute and chronic harms, including the erosion of trust in authoritative sources and democratic institutions over time.
- Finding 9:** The real-world consequences of misinformation have been readily apparent over the past 18 months: inciting violence, undermining official health advice, and causing tangible financial impacts on governments, industry and consumers.
- Finding 10:** Most Australians are aware of platform measures to remove or label offending content, but few have direct experience. Early evidence suggests these steps have been somewhat effective in reducing amplification of misinformation on particular platforms.
- Finding 11:** Australians see the issue of misinformation to be one of joint responsibility – split between individual users, platforms, and government. There is some scepticism in the ability of platforms to self-regulate, and concern about government’s role in regulating speech.
- Finding 12:** Information on the effectiveness of platform measures is limited, and more needs to be done to better understand what measures work and to monitor the effectiveness of platform moderation activities.
- Finding 13:** In leading code development, DIGI successfully managed a novel, complex and time-sensitive project, navigating a range of competing interests across a disparate group of stakeholders that included both members and non-members of DIGI.
- Finding 14:** DIGI undertook a meaningful public consultation process on its draft code, generating a range of feedback from academia, industry, and parts of civil society, which visibly informed the final code.
- Finding 15:** DIGI could have improved its consultation process with greater publicity, including promoting it through existing public communications channels and engagement with the media.
- Finding 16:** DIGI dealt with stakeholder feedback in a relatively open and transparent manner. However, the significant change in scope meant it would have been best practice to provide stakeholders a further opportunity to comment on the final drafting prior to finalisation.
- Finding 17:** The bulk of ‘major platforms’ in Australia have signed up to the code. As such, it should be regarded as an industry-wide initiative.
- Finding 18:** DIGI should continue to encourage other popular platforms, like Snapchat and Reddit, to sign up to the code, even if they do not meet the proposed threshold of one million active monthly users. DIGI should actively publicise the involvement of any additional code signatories as soon as practicable after their signing.
- Finding 19:** Industry participants should consider the role of private messaging platforms and smaller alternative platforms in the amplification of disinformation and misinformation and explore options for how these platforms could be included within the code framework.

- Finding 20:** DIGI has developed an outcomes-based code that has allowed platforms with a range of business models to sign up to a single code.
- Finding 21:** The code objectives and principles meet the government objective of striking a balance between encouraging platform interventions and protecting freedom of expression, privacy and other rights
- Finding 22:** The code should be strengthened by taking an opt-out approach. Opting out of an outcome should be permitted only where the outcome is not relevant to the signatory's services. Signatories should provide adequate justification when opting out.
- Finding 23:** The code covers both disinformation and misinformation. This is one of the key strengths of the code, and is an improvement on the current EU Code.
- Finding 24:** The definition of harm in the code is too narrow to provide adequate safeguards against the full range of harms caused by the propagation of disinformation and misinformation.
- Finding 25:** Private messaging services should be included within the scope of the code as these are known vectors of disinformation and misinformation. These should be included with appropriate caveats on the right to privacy.
- Finding 26:** The code should clarify that the exclusion of professional news content applies only to the application of counter-misinformation measures. It should also clarify that news aggregation services are in scope.
- Finding 27:** The treatment of paid and sponsored content should be clearer in the code. This should include a clear definition of sponsored content.
- Finding 28:** In addition to improving public awareness of the source of political advertising, the code should also cover the source of issues-based advertising.
- Finding 29:** The output-focused framing of several objectives and measures may encourage reporting to focus on outputs rather than progress towards code outcomes.
- Finding 30:** The code should include industry-wide frameworks for the development and implementation of individual platform measures. Examples could include frameworks to establish:
- > criteria for the assessment of harm
 - > criteria for assessing news and information quality
 - > processes for the exchange of information between platforms on disinformation and misinformation risks
 - > commitments to address the propagation of disinformation and misinformation via platform advertising channels
 - > commitments to address the risks of propagation via platform algorithms and architecture.
- Finding 31:** The code provides a high-level code administration framework. Given that detailed arrangements for code administration, compliance with the code, and consumer complaints are still under development, the ACMA's ability to assess their practical effectiveness is constrained.

- Finding 32:** The code should include a framework setting out principles for the structure and operation of the sub-committee to provide greater transparency and accountability.
- Finding 33:** The reporting template provides a workable foundation for the reporting guideline. Reporting should incorporate adequate data to measures performance against KPIs under each outcome; detailed action plans to address areas identified for improvement; and a clearer distinction between measures (that is, outputs) and the effectiveness of these measures (progress towards outcomes).
- Finding 34:** The proposed 12-month review will provide an opportunity for findings from this report, and other developments, to be incorporated into the code.
- Finding 35:** The lack of detail on code administration matters, including on the operation of the sub-committee and guidelines for future code reporting, has limited the ACMA's ability to undertake a full assessment on the likely effectiveness of the code.
- Finding 36:** On the whole, signatories have met the initial reporting requirements set out in the code.
- Finding 37:** For the most part, signatories have provided appropriate explanations where they have not opted-in to specific commitments.
- Finding 38:** A more uniform approach to reporting would assist in cross-platform assessment and increase transparency of platform measures and performance.
- Finding 39:** For future reports, signatories should clearly specify the products and services covered by the code, and justify any major exclusions.
- Finding 40:** Signatories have a wide range of measures in place to address the problems of disinformation and misinformation and to improve the quality of news and information on their services. They also demonstrate responsiveness to significant changes over the last 18 months, as well as to public and government calls for stronger action.
- Finding 41:** It is expected that signatories will develop more Australia-focused measures over time.
- Finding 42:** In general, the reports are heavily focused on current measures and past actions, and signatories have provided little systematic information on future initiatives. In some cases, it is not clear to what extent certain measures will contribute to the achievement of the code outcomes under which they have been reported.
- Finding 43:** There are inconsistencies in the interpretations of key terms between signatories, which are drawn from pre-existing definitions from their internal, often global, policies. This makes it difficult to interpret and assess performance and to make industry-wide comparisons.
- Finding 44:** A harmonised template would assist in comparing initiatives across platforms. It would also allow clear reporting on additional information beyond the requirements of the code.

- Finding 45:** Signatories have provided a large range of information on the actions they have taken to address misinformation, disinformation and news quality and to invest in collaborative initiatives. This demonstrates signatories' commitment to addressing these issues.
- Finding 46:** The information signatories have provided is heavily focused on platform outputs and on volumetric data. Reporting lacks systematic data, metrics or key performance indicators (KPIs) that establish a baseline and enable the tracking of platform and industry performance against code outcomes over time.
- Finding 47:** Reports provide some data on the Australian context, but this is often piecemeal or not directly related to actions under the code. Reporting should include Australia-specific data and signatories should establish a reporting regime against the Australian code.
- Finding 48:** Reporting lacks trend-related data. Trend-related data would contribute to a greater understanding of the extent and impact of disinformation and misinformation in Australia.

Appendix B: Signatory assessment reports

 Adobe	
Australian user base	Unknown
Services covered:	All Adobe services able to take advantage of the Content Authenticity Initiative (CAI), including Creative Cloud
Code commitments:	5.8, 5.9, 5.18, 5.19, 5.28, 5.29, 5.30
Strengths <ul style="list-style-type: none"> > Adobe has signed up to the code based on the CAI, opting into the outcomes that it considers proportionate and relevant to its business. > The report outlines the importance of the CAI and associated C2PA standards to improve attribution of digital content, both for users of Adobe products, and more broadly across industry. The level of information provided in the report is appropriate given Adobe’s limited ranges of commitments under the code. > The inclusion of software providers, like Adobe, strengthens the code by broadening its remit to include those who can provide tools and solutions for addressing disinformation and misinformation. This may encourage greater collaboration on common solutions between platforms and providers. 	
Weaknesses <ul style="list-style-type: none"> > No Australia-specific measures are identified in the code, beyond briefings and events to promote consideration of content authenticity and increase understanding of attribution. Similarly, Adobe provides no Australia-specific data. > While Adobe has a goal for the future of the CAI, it doesn’t provide any concrete data, discuss targets, or identify what metrics it will use to measure success. > Adobe provides an overarching rationale for its limited involvement, but it remains unclear why it has chosen not to opt-in to some of the code outcomes that appear relevant to the CAI, such as developing partnerships and collaborating on strategic research (Outcome 6). 	
Recommendations <p>Adobe should consider publishing information on its measures of success under the code, such as targets for the uptake of CAI in Australia. Adobe should also consider whether it could opt-in to other code outcomes that appear relevant to the CAI.</p>	



Apple

Australian user base	Unknown
Services covered:	Apple News
Code commitments:	1a, 1d, 2, 4, 6, 7 (5.28 only)
General comment Apple's commitments under the code are limited to its Apple News service. While we welcome Apple's involvement, certain aspects of its commitment remain unclear, due to the code's treatment of news aggregation services and the exclusion of professional news content from the operation of the code.	
Strengths <ul style="list-style-type: none">> Apple has opted-in to several commitments in relation to Apple News. For most of the provisions it has opted-out of, it has provided a short explanation for why these provisions are not applicable to Apple News.> Apple's report is detailed and contains several Australian-specific examples and commitments. These include its approach to the 2019-20 Australian bushfire season and the employment of Australian journalists and editors at Apple News.	
Weaknesses <ul style="list-style-type: none">> It is unclear whether some measures in the report are existing or future initiatives.> The report provides minimal data on Apple's content moderation activities. The one data point on content removal is not limited to disinformation and misinformation, and it is a global statistic.> Apple did not include a reason why it has chosen not to opt-in to all measures under Outcome 7, including publishing additional information detailing its progress against implementing code commitments.	
Recommendations Apple should consider providing more Australian-specific data to demonstrate how its relevant policies and initiatives are helping address the issues of disinformation and misinformation.	

FACEBOOK

Australian user base:	17 million monthly active users (Facebook), 9 million monthly active users (Instagram)
Services covered:	Facebook, Instagram
Code commitments:	All code outcomes

Strengths

- > Facebook sets out 43 specific commitments under the 7 code outcomes. Commitments include 4 new Australian initiatives – 2 due to be rolled out in 2021. The expansion of its political-advertising policy to social-issue advertisements commenced on 29 June 2021.
- > The report provided detailed explanations of Facebook’s global policies and initiatives for combatting misinformation and disinformation, including:
 - > coordinated inauthentic behaviour (disinformation) policies
 - > a graduated approach to harmful misinformation which has adjusted in response to more recent developments
 - > measures to promote authoritative information and provide users with tools to help them assess the quality of sources and factual content.
- > It also includes Australia-specific data and information on its recent actions relating to COVID-19 misinformation, and information on its investment in research and collaborative initiatives including funding academic research, supporting local events and developing misinformation-focused training.
- > Facebook was the only signatory to broadly publicise the release of its report.

Weaknesses

- > Reported data is piecemeal, and Facebook does not establish any metrics that could be used to track the effectiveness of its measures, particularly in the Australian context. It is noted that this will be developed over time.
- > While Facebook provides some information on its WhatsApp and Messenger measures, it does not consider these services to be covered under the code. It is unclear whether Facebook will continue reporting on these services.
- > The report adopts internal definitions, with disinformation being used to refer to inauthentic behaviour with the intention to deceive, and misinformation to refer to content that is false or misleading. This is inconsistent with the code.

Recommendations

Facebook should provide more-extensive data against robust KPIs that can be used to assess the effectiveness of its measures. Although private messaging services are currently excluded, Facebook could consider nominating its messaging services under the code to provide greater transparency over its existing activities.



Australian user base	18 million unique monthly users (Google Search), 16.5 million unique monthly visitors (YouTube)
Services covered:	Google Search, YouTube, Google Ads
Code commitments:	All code outcomes

Strengths


- > Google signed up to all outcomes under the code and provided a detailed initial report highlighting a range of relevant policies and measures.
- > The report outlines the categories of intervention, its relevant global policies, and the mechanisms for users to report violations against these policies.
- > Google identifies several Australia-specific measures, such as new policies and identity verification requirements for Australian election ads, and funding of Australian research, including through the Google News Initiative.
- > The report provided a detailed case study on steps taken to address COVID-19 misinformation, including promoting official health advice from the Australian Government, providing ad grants to the government and local NGOs, and partnering with local YouTube creators on public service announcements.
- > Google noted that it views its commitments under the code as a floor rather than a ceiling and will strive for continual improvement over time.


Weaknesses


- > There is no identification or discussion of future measures, or what continual improvement could look like in practice under the code.
- > Google provides limited evidence demonstrating the effectiveness of its existing measures, or no data on trends relating to disinformation or misinformation over time.
- > The reports don't provide a rationale as to why some but not other Google services are out of scope. The report includes some information on measures relating to Google News but it isn't clear if Google News is covered under the code.
- > Google published limited data directly related to Australia, despite comparable breakdowns being publicly available for other countries (e.g., via YouTube Community Guidelines enforcement).

Recommendations

For future reporting, Google should seek to provide further detail about how it measures success, publicly report on more Australia-specific data, and discuss future initiatives or planned improvements under the code.

 Microsoft	
Australian user base	6.5 million monthly active users (LinkedIn), 1.7 million monthly active users (Bing)
Services covered:	All consumer-facing products
Code commitments:	All except 5.28 (undecided)
Strengths <ul style="list-style-type: none"> > Microsoft has committed to nearly all outcomes under the code, across all of its consumer-facing products. > Microsoft provides a broad list of relevant programs and initiatives that it has undertaken to address the issues of disinformation and misinformation. > The report includes some recent actions and data relating to COVID-19 advertising, and details investment into emerging issues, such as data voids and deepfakes. 	
Weaknesses <ul style="list-style-type: none"> > It is not clear which Microsoft services are covered by the code. Microsoft has said its report covers its 'consumer-facing services' (referencing Bing, LinkedIn, Microsoft News, and Microsoft Advertising), however the extent of this service-type is nebulous. While the ACMA appreciates this could possibly extend to a growing list of hundreds of services, a more clear and specific scope or list of services covered would be beneficial. > Data does not establish any metrics that could be used to track the effectiveness of its measures. Of the minimal data provided, it is not always used to show how effective a relevant policy or initiative is. The published report also does not provide any Australia-specific data points. > While Microsoft refers to measure 5.27 in its annual report, its commitment is left blank in the nomination form. In discussions with the ACMA, Microsoft said it has yet to decide whether the opt-in to 5.27 and is awaiting further detail from DIGI. This could have been made clear in its nomination form. > The report references Microsoft's 'On the Issues' blog as a place where company announcements about technology policy issues, including those relating to disinformation and misinformation, are hosted. There is currently no reference to the Australian code on this site. 	
Recommendations <p>To improve transparency, Microsoft should specify the services it considers to be covered by the code. Microsoft should also consider providing more Australia-specific data to build KPIs and demonstrate how its relevant policies and initiatives are helping address the issues of disinformation and misinformation domestically.</p>	

 REDBUBBLE	
Australian user base	Unknown
Services covered:	Redbubble
Code commitments:	All except 1(d), 4, 5 and part of 2. As an online marketplace, Redbubble has opted out of Outcomes 4 and 5 as it does not provide news, factual content or third-party advertising.
<p>Strengths</p> <ul style="list-style-type: none"> > The report is clear and consumer-focused, and Redbubble provides clear and reasonable justifications for opting out of certain code outcomes. > Redbubble provides trend-related data on misinformation on its platform, including anti-vaccination content, and a general summary of future initiatives relating to misinformation and disinformation. > The report provides a clear explanation of relevant policies and how these are implemented. > Redbubble uses credible and trusted sources, including independent fact checking sites, to determine the boundaries of disinformation and misinformation and compile guidelines. Precedent and global context are also used to inform decisions and policies. 	
<p>Weaknesses</p> <ul style="list-style-type: none"> > Redbubble has not opted into Outcome 1(d) (provide general information on actions taken in response to user reports), for the reason that reporting is currently anonymous and no mechanism for publishing Redbubble's responses is currently available. > Redbubble provides only a general commitment to research and collaboration under Outcome 6 and no further commitment to transparency under Objective 7 beyond the code report. However, Redbubble has a small Australian user base and limited exposure to disinformation and misinformation. > Redbubble's policies include a definition of harmful misinformation, but do not refer to disinformation as defined in the DIGI code. However, Redbubble's definition of misinformation appears to be broad enough to cover both concepts and their report includes information on its measures relating to demonetisation and service integrity under Outcomes 2 and 3 respectively. 	
<p>Recommendations</p> <p>As Redbubble has a user-reporting function for reporting misinformation, it should consider opting into Outcome 1(d) and working towards the provision of suitable transparency mechanisms under the code. Redbubble could also consider involvement in more formal research and collaboration partnerships to feed into its content policy frameworks. To further strengthen its future reporting, Redbubble could consider including data on the actions it has taken against disinformation and misinformation content.</p>	

	
Australian user base	1.8 million monthly active users
Services covered:	TikTok
Code commitments:	All code outcomes
<p>Strengths</p> <ul style="list-style-type: none"> > As a smaller platform and more recent entrant to the Australian market, it is pleasing to see TikTok sign up to all code outcomes > TikTok has provided some quantitative Australian-specific data on COVID-19 and medical misinformation video removals, and numbers of unsubstantiated COVID-19 claims tagged with information notices relevant to Outcome 1a. > The report includes month-by-month breakdown of data, allowing for a deeper insight into the effectiveness of some of its policies over time. > The report followed the agreed reporting template, providing sufficient contextual information about its approach to disinformation and misinformation, before reporting on specific against code outcomes. 	
<p>Weaknesses</p> <ul style="list-style-type: none"> > TikTok's report does not fully align with the definitions of disinformation and misinformation used in the code. TikTok's definition of misinformation, for example, does not capture misleading information, except to the extent that it relates to elections or civic processes and synthetic or manipulated media. > TikTok does not clearly articulate its approach to assessing harm and deciding whether content should be labelled or removed. > It is not clear how TikTok's reported measure against Outcome 6 (Asia-Pacific Advisory Council) will contribute to the achievement of the code outcome. 	
<p>Recommendations</p> <p>TikTok should provide greater detail how its policies and broader measures align to the code outcomes, and better signal what future initiatives it is considering, particularly in relation to Australian users.</p>	



Twitter

Australian user base	5.8 million monthly active users
Services covered:	Twitter
Code commitments:	All code outcomes, except Outcome 5

Strengths

- > Twitter has opted into all code outcomes except Outcome 5 on the basis that it does not accept political advertising.
- > The report shows meaningful commitment to action against misinformation in the context of specific issues such as COVID-19 and elections, and more general commitment surrounding disinformation, information quality, and public transparency. Twitter provides some detailed information on its policies and tools on disinformation and misinformation under Outcome 1.
- > Twitter provides some specific data on its actions against misinformation and disinformation, including account suspensions, removals and content removals, both globally and for Australian accounts.
- > COVID-19 policy evidences a responsive, proportionate, risk-based approach.

Weaknesses

- > The report does not address how the problems of disinformation and misinformation manifest on the platform, with no data on trends or prevalence that would facilitate benchmarking.
- > While Twitter has shared some information with the ACMA on a confidential basis, it would be beneficial for more information on future measures in response to the code or broader environmental developments, either globally or in Australia were included in the report.
- > The data provided is not sufficiently comprehensive or specific to Australia. For example, the report gives no data on actions taken against synthetic and manipulated media and provides only global numbers on content labelling for misinformation.
- > Does not address how the principles of the COVID-19 misleading information or civic integrity policies might be applied to other areas of potential harm, or whether their user-reporting mechanism for manipulated media will be extended to other areas of disinformation and misinformation in accordance with Outcome 1c.

Recommendations

Twitter should strengthen its reporting to include KPIs that track performance under the code, detailed Australia-specific data, trends, and planned measures. Twitter should consider expanding the risk-based principles in its COVID-19 misinformation policy to cover misinformation more generally and its user-reporting mechanism to cover the scope of misinformation and disinformation set out in the code.






Appendix C: Timeline of key events












Since the government published its response to the DPI in December 2019, digital platforms have implemented a range of changes to address disinformation and misinformation on their platforms. The following timeline traces these activities against key social and political events that have impacted disinformation and misinformation trends. Platform responses include policy changes, enforcement actions, changes to the platform features and functions, and other actions and initiatives.

Almost all policy changes and new initiatives identified by the ACMA have been in response to either the COVID-19 pandemic or the 2020 US presidential election.

In high-level summary, platforms have:

1. invested in third-party fact-checking organisations to proactively identify and flag false information on their service
2. proactively updated their policies to specifically address unique events where there is a heightened risk of harm and increased enforcement actions against potentially misleading and false information in relation to unique events
3. invested in technological means to signal credible, relevant and authentic information
4. signalled information that may be false or misleading and provided additional context by linking to authoritative information sources, such as those published by World Health Organisation (WHO)
5. provided financial assistance and grants to news outlets and government and not-for-profit organisations to bolster the spread of credible information and news
6. increased detection, monitoring and enforcement action against groups and networks who use their services to spread disinformation and misinformation.

Timeline key	
	Significant social or political events
	Policy changes
	Enforcement actions
	Changes to the platform features and functions
	Other actions or initiatives.
Australian events or platforms measures directly relevant to Australia	

December 2019	
12 Dec	 Australia – Australian Government responds to the ACCC’s Digital Platforms Inquiry; calls on digital platforms to establish a voluntary disinformation code.
31 Dec	 China – First cases of a novel coronavirus reported in Wuhan.
January 2020	
Early Jan	 Australia – South-eastern bushfires (‘Black Summer’) at peak. Fires in NSW and Victoria are extinguished or contained by early March.
6 Jan	 Facebook announces changes to its Manipulated Media Policy to remove videos that are the product of machine learning techniques (e.g., deepfakes) and have been edited or synthesised in a manner likely mislead an average person to believe a subject of the video said words they did not say. ¹¹⁹
8 Jan	 TikTok publishes expanded and more detailed Community Guidelines intended to be easier to understand and enforce. ¹²⁰ The guidelines prohibit content intended to mislead, including synthetic content (e.g., shallow or deep-fakes), spam and disinformation. ¹²¹ Misinformation that may cause harm, including harm to an individual’s health, or that may mislead the public about elections and civic processes, is also prohibited. ¹²²
13 Jan	 Thailand – First cases of coronavirus outside China reported.
25 Jan	 Australia – First local case of coronavirus reported.
29 Jan	 Canada – <i>Canada’s Communications Future: Time to Act</i> report recommends legislative reform to address harmful content on digital platforms.
	 Twitter launches a new search prompt, in partnership with authoritative health agencies (including with the Australian Department of Health) to help surface credible COVID-19 information and remove auto-suggest options likely to lead to non-credible information. ¹²³
30 Jan	 Global – WHO declares the COVID-19 outbreak a Public Health Emergency of International Concern.
Late Jan	 Facebook announces it will apply its existing Misinformation and Harm policy (covering Facebook and Instagram) to remove COVID-19 misinformation that could cause physical harm. For example, claims that drinking bleach can cure the virus, or that social distancing is ineffective. Facebook has had policies in place to remove misinformation that can lead to serious and imminent physical harm since January 2018. ¹²⁴

¹¹⁹ Monika Bickert, [Enforcing Against Manipulated Media - About Facebook](#).

¹²⁰ TikTok [submission to the Senate Select Committee on Foreign interference through social media](#).

¹²¹ TikTok, [Community Guidelines](#).

¹²² TikTok, [Building to support content, account, and platform integrity](#).

¹²³ Jun Chu and Jennifer McDonald, [Helping the world find credible information about novel #coronavirus](#).

¹²⁴ Facebook’s [Submission to the Senate Select Committee on Foreign Interference through Social Media](#),

p 9.



Facebook updates its policies to limit the spread of information that would otherwise not contravene its Community Standards, but that promote misleading information about the pandemic or discourages vaccination.¹²⁵ It will remove false information that falls into a number of broad categories of COVID-19 related misinformation: false information on the existence or severity of COVID-19, links to 5G technologies, transmission and immunity, cures or prevention methods and information that discourages good health practice.



Google treats the COVID-19 pandemic as a ‘sensitive event’ for the purposes of applying its policies, which variously prohibit behaviours and content-related to disinformation and misinformation.¹²⁶



TikTok makes a number of specific updates to its policies in response to the COVID-19 pandemic, more readily remove or limit COVID-19 misinformation, including to:¹²⁷

- > remove false medical advice that may cause harm
- > remove false information that is likely to cause panic
- > limit the dissemination of false claims that the pandemic was deliberately spread
- > remove content that suggests certain groups of people are more likely to have or spread COVID-19.



TikTok places restrictions on advertising on its platform, including a prohibition against advertising that advocates against vaccination.¹²⁸



Redbubble publishes guidance on coronavirus/COVID-19 content. The guidance sets out certain types of COVID-19 related content that may be removed from the site, including:

- > messages which discriminate against certain groups
- > jokes at the expense of victims
- > works with ambiguous or harmful intent
- > unnecessarily graphic content
- > messages spreading false information or causing panic.¹²⁹

February 2020



Facebook conducts a ‘Strategic Network Disruption’ against the Proud Boys.¹³⁰ This involves removing individual and group accounts to prevent extremist groups from using Facebook as an organising tool. Facebook repeats this exercise in June and September 2020.



Apple News launches special coverage of the 2020 US presidential election for US users. It includes a ‘news literacy guide’ developed in partnership with the News Literacy Project, to help users identify misinformation.¹³¹

1 Feb



UK – EU–UK Withdrawal Agreement comes into force.

¹²⁵ Instagram, [COVID-19 and Vaccine Policy Updates and Protections](#).

¹²⁶ Google, [Google Initial transparency report, May 2021](#).


¹²⁷ TikTok, [submission to the Senate Select Committee on Foreign Interference Through Social Media](#).


¹²⁸ TikTok, [Supporting our community through COVID-19](#).

¹²⁹ Redbubble, [Coronavirus/COVID-19 Content on Redbubble](#).


¹³⁰ Facebook, [Taking action to combat misinformation, polarization, and dangerous organizations](#).


¹³¹ Apple Newsroom, [‘Apple News launches special coverage of the 2020 presidential election’](#).


4 Feb  **Twitter** publishes its policy on synthetic and manipulated media.¹³² The policy indicates this content is prohibited and subject to a range of enforcement actions, such as labelling.¹³³ Synthetic or manipulated media that is likely to cause harm, such as physical harm or civil unrest, will be removed. Users may report content they believe may contravene this policy.


24 Feb  **Instagram** announces new measures to support accurate COVID-19 information, including educational resources and limiting COVID-19 accounts accessible in recommendations to credible health organisations.¹³⁴


March 2020


 **YouTube** introduces COVID-19 information panels on its home page and on pages related to the pandemic.¹³⁵


 **Apple** is reported to only be approving apps related to COVID-19 if they are developed by official institutions.¹³⁶ Apple reviews apps for release on its app store under its App Store Review Guidelines.¹³⁷


 **TikTok** opens a US-based Transparency Centre, which will enable independent experts to observe how TikTok implements its content moderation and complaints-handling policies.¹³⁸


3 Mar  **Facebook** implements a new policy to prevent advertisers from exploiting the pandemic for financial gain and introduces information pop-ups on top of search results directing users to the WHO and local health authorities to assist users to find reliable information on the COVID-19 pandemic.

5 Mar  **TikTok** launches a COVID-19 Information Hub in the UK. Users who explore hashtags related to COVID-19 pages are directed to the Hub, which includes information from trusted sources, such as the WHO and the Red Cross.¹³⁹

11 Mar  **Global** – WHO declares COVID-19 a pandemic.

13 Mar  **Australia** – National Cabinet formed in response to the COVID-19 pandemic.

17 Mar  **Facebook, Google, LinkedIn, Microsoft, Twitter, YouTube and Reddit** issue a joint statement outlining their approach to addressing misinformation about COVID-19 on their respective platforms.

17 Mar  **Facebook** launches a US\$1 million grant program with The International Fact-Checking Network and a further US\$1 million worth of grants to local news

¹³² Twitter, [Synthetic and manipulated media policy](#).

¹³³ Achuthan, A. and Roth, Y. [Building rules in public: Our approach to synthetic & manipulated media](#).

¹³⁴ Instagram, [Keeping people safe, informed and supported on Instagram](#).

¹³⁵ Google, [How you'll find accurate and timely information on COVID-19 vaccines](#).


¹³⁶ Kif Leswing, ['Apple rejecting coronavirus apps that aren't from health organisations, app makers say'](#), CNBC.


¹³⁷ Apple, [App Store Review Guidelines](#).

¹³⁸ Sarah Perez, ['TikTok to open a 'Transparency Centre' where outside experts can examine its moderation practices'](#), Techcrunch.

¹³⁹ TikTok [submission to the Senate Select Committee on Foreign Interference Through Social Media](#).


organisations to support costs related to covering the COVID-19 pandemic.¹⁴⁰ If a piece of content is rated false by fact-checkers, Facebook will limit its distribution and label it with a warning and information to provide more context. Users are unable to view such content without clicking through the warning label. Facebook states that in 95% of cases, users will not proceed to view the false content.¹⁴¹ Facebook states that: ‘in April 2020 alone, we applied the label and reduced the distribution of more than 50 million posts worldwide, based on more than 7,500 fact-checks’.¹⁴²


18 Mar  **Facebook** launches the Facebook Coronavirus Information Center, providing a centralised location for users to access news and information about COVID-19. In Australia, the Center includes official Australian Government information.


18 Mar  **Twitter** broadens its the definition of harm to address content that contradicts guidance on COVID-19 from ‘authoritative sources of global and local public health information’.¹⁴³ Under Twitter’s COVID-19 misleading information policy, content may be labelled or removed if it definitively advances a false claim, is demonstrably misleading and likely to cause serious harm.¹⁴⁴ There are 5 categories of information covered by the policy, include false or misleading information about:

- > the nature of the virus
- > the efficacy and/or safety of preventative measures, treatments, or other precautions to mitigate or treat the disease
- > official regulations, restrictions, or exemptions pertaining to health advisories
- > the prevalence of the virus, or risk of infection or death.

Twitter provides local authorities and experts with access to a Partner Support Portal to assess information and escalate problematic content for human review.¹⁴⁵ Content may be removed under this policy and repeated contraventions may result in an account being permanently suspended. The policy also makes exemptions for commentary, satire, counter speech, first-hand accounts or public debate for the purposes of advancing COVID-19 science and research.

20 Mar  **Australia** – National borders closed to non-residents.

20 Mar  **US** – President Trump claims that hydroxychloroquine could prevent or treat COVID-19.

20 Mar  **Facebook** launches the WHO Health Alert on **WhatsApp**. The alert provides reliable information on matters such as how to prevent the spread of COVID-19 and correcting false information and conspiracy theories. The alert

¹⁴⁰ Facebook, [Keeping People Safe and Informed About the Coronavirus](#).

¹⁴¹ Facebook, [submission to the Senate Select Committee on Foreign Interference through Social Media](#), p 10.


¹⁴² Facebook, *Facebook response to the Australian disinformation and misinformation industry code*, p. 19.


¹⁴³ @Vijaya and Matt Derella, [An update on our continuity strategy during COVID-19](#).


¹⁴⁴ Twitter, [COVID-19 misleading information policy](#).

¹⁴⁵ Twitter, *Australian code of practice on disinformation ad misinformation initial report*, p.14.


is first made available in English, and then rolled out in Arabic, Chinese, French, Russian and Spanish.¹⁴⁶

- 24 Mar  **Facebook** announces a range of updates to improve the accuracy of COVID-19 information on **Instagram**, including:
- > educational resources
 - > labels on accurate authoritative information
 - > only recommending COVID-19 accounts from credible health organisations.¹⁴⁷


29 Mar  **Australia** – COVID-19 social distancing rules strengthened, establishing 2-person limit for all gatherings, and new lockdown restrictions.


30 Mar  **UK** – The Cabinet Office and Number 10 begins implementing a specialist rapid response unit to combat COVID-19 misinformation.


April 2020

2 Apr  **Google** announces a range of new funding and initiatives intended to improve COVID-19 news quality.¹⁴⁸ The Google News Initiative provides additional support to First Draft for an online resource hub for journalists and to the Crosscheck Network, which helps newsrooms respond to escalating harmful content. In Australia, Google provides funding for media literacy research and educational resources and works with the ABC to launch a COVID-19 news bulletin on Google Assistant.¹⁴⁹

Google also gives US\$5 million to support fact-checking programs,¹⁵⁰ particularly in Europe (focusing on the countries with the highest number of COVID-19 cases) and in Spanish-speaking and Latin-American countries.

6 Apr  **Google** implements new policies to mitigate misleading claims and promote transparency on COVID-19 related information provided in apps available through Google Play.¹⁵¹ Only official government apps and verified health apps which provide medical or support services related to COVID-19 or which support local responses, may leverage COVID-19 related keyword searches in the Google Play Store.

7 Apr  **Facebook** limits forwarded messages in **WhatsApp** to reduce viral misinformation. Once a message has been forwarded through a chain of 5 or more chats, the message is labelled with a double arrow icon and may only be forwarded to a single person at a time.¹⁵²

14 Apr  **Global** – United Nations Communications Response initiative to combat the spread of disinformation and misinformation is launched.

¹⁴⁶ Facebook, [Keeping People Safe and Informed About the Coronavirus](#).

¹⁴⁷ Instagram, [Keeping people safe, informed and supported on Instagram](#).



¹⁴⁸ Google, [\\$6.5 million to help fight coronavirus misinformation](#).


¹⁴⁹ Google, *Australian code of practice on disinformation and misinformation: Google Initial Report, May 2021*.

¹⁵⁰ Google, [How you'll find accurate and timely information on COVID-19 vaccines](#).


¹⁵¹ Sam Tolomei, '[Google Play updates and information: Resources for developers](#)', *Android Developers Blog*.


¹⁵² Casey Newton, [WhatsApp puts new limits on the forwarding of viral messages](#).


- 21 Apr  **Twitter**, in partnership with UNESCO, launches the #ThinkBeforeSharing campaign, intended to raise awareness about conspiracy theories and improve media literacy.¹⁵³
- 22 Apr  **Facebook** announces modifications to provide location information for posts shared by high-reach pages, to increase users' ability to discern the 'reliability and authenticity' of posts they see in their feeds.


28 Apr  **Republic of Ireland** – BAI publishes *CodeCheck: A Review of Platform Compliance with the EC Code of Practice on Disinformation*.


May 2020


6 May  **Facebook** announces the membership of its Oversight Board.¹⁵⁴ The Board was formed to make final and binding decision on whether specific content should be allowed or removed from Facebook and Instagram. Users are able to appeal Facebook moderation decisions to the Board to have their content restored. Cases may also be referred to the Board directly by Facebook.

9 May  **Australia** – Anti-lockdown protests in Melbourne.

11 May  **Twitter** introduces additional labels and warning messages on tweets about COVID-19 containing misleading information or disputed claims. Labels are applied retrospectively, where false and misleading content has a 'moderate' propensity for harm. Labelled content is identified using proactive 'internal systems' and via fact-checking partnerships.¹⁵⁵

19 May  **Global** – WHO Member States pass Resolution WHA73.1, which recognises that managing the 'infodemic' is critical to controlling the COVID-19 pandemic.

20 May  **YouTube** publishes a COVID-19 Medical Misinformation Policy.¹⁵⁶ The policy prohibits content that contradicts guidance by the WHO or local health authorities on the treatment, prevention, diagnosis, transmission or existence of COVID-19 or social distancing and self-isolation advice/rules. An exception to the policy is provided for educational, documentary, scientific or artistic content that contextualises such claims as false. This policy is subsequently updated on 14 October 2020 to specifically address vaccines. Content that contravenes the policy is removed and, after 3 contraventions, offending channels are terminated. Google stated that this policy was developed out of its ongoing work with global and local health authorities to ensure its policies are accurately addressing risks of serious physical harm or death.¹⁵⁷

22 May  **Facebook** makes location information available on posts shared by high-reach pages.












¹⁵³ UNESCO, [European social media campaign to address disinformation on Covid-19 & #ThinkBeforeSharing](#).

¹⁵⁴ Nick Clegg, [Welcoming the Oversight Board](#).

¹⁵⁵ Yoel Roth and Nick Pickles, [Updating our approach to misleading information](#).

¹⁵⁶ YouTube, [COVID-19 medical misinformation policy](#).

¹⁵⁷ Google, [Australian code of practice on disinformation and misinformation: Google Initial Report](#), May 2021.

26 May		US - Black Lives Matter protests commence, following the killing of George Floyd in Minneapolis. Protests occur in more than 60 countries over the coming weeks.
28 May		US – President Trump signs the Executive Order on Preventing Online Censorship, seeking reform of online platform protections.
28 May		Facebook expands its policy of verifying the identity of users managing pages with large audiences to also verify accounts generating viral posts in the US that 'have a pattern of inauthentic behaviour'. ¹⁵⁸ The purpose of the policy change was to provide additional transparency and information to Facebook users. If an account chooses not to confirm its information, its content may be disabled or have diminished reach.
30 May		Australia – Anti-vax and anti-5G protests held in Sydney, Brisbane and Melbourne.
June 2020		
		Facebook implements controls to enable US Facebook and Instagram users to limit electoral, political and social issue advertising. Facebook already requires that advertising for electoral, political and social issues (sensitive topics that are heavily debated, may influence the outcome of an election or result in/relate to existing or proposed legislation) are authorised and include 'paid for by' disclaimers. Facebook states it will be expanded to social issue advertising in 2021. ¹⁵⁹ An archive of political advertisements can be searched publicly in Facebook's Ad Library.
1 Jun		Australia – NSW Court of Appeal upholds 2019 decision, which found news media companies liable for defamatory comments posted by users on their public Facebook pages.
1 Jun		Twitter discloses 3 networks of accounts to its archive of state-linked operations.
4 Jun		Facebook commences labelling of content from state-controlled media outlets. On 17 June, Facebook also starts blocks advertising from state-controlled media outlets targeting people in the US.
6 Jun		Australia – Black Lives Matter protests in Sydney.
20 Jun		Australia – Victoria experiences COVID-19 second wave, and reintroduction of state-wide lockdown restrictions. Restrictions ease on 22 November.
25 Jun		Facebook introduces a new notification to alert users if a news article they are going to share is more than 90 days old. The intention being to improve context, in addition to the existing feature 'context button', which gives users information on the source of an article. ¹⁶⁰


¹⁵⁸ Anita Joseph, [Verifying the identity of people behind high-reach profiles](#).


¹⁵⁹ Facebook, *Facebook response to the Australian disinformation ad misinformation industry code*, p. 8.


¹⁶⁰ Facebook, [Providing People With Additional Context About Content They Share](#).


26 Jun ○○○ **Australia** – ACMA publishes *Misinformation and news quality on digital platforms in Australia: A position paper to guide code development*.¹⁶¹

July 2020


15 Jul  **Facebook** updates its Facebook Coronavirus Information Center to include a dedicated section on common myths (identified by the WHO) on COVID-19.

29 Jul  **US** – ‘Anti-trust hearing’: Apple, Google, Facebook and Amazon appear before US House Judiciary Committee.

31 Jul  **Google** makes changes to its advertising policies, prohibiting websites that accept Google advertising from publishing hacked materials, and to prohibit advertisements that directly facilitate or advertise access to hacked material related to political entities, including those distributed by a third party.¹⁶² This policy is rolled out globally in November 2020.

31 Jul  **Google** announces that it will implement a new feature called ‘About this Ad’, which will enable users to see the verified name of an advertiser.¹⁶³ Additionally, Google announces the release of a new tool called Ads Transparency Spotlight (an alpha extension from the Chrome Web Store), which provides detailed information about all advertising on Google Search.

August 2020

 **Facebook** implements a number of changes to diminish the spread of harmful content and extremism on its platforms, including:¹⁶⁴

- > instituting a waiting period to prevent newly created Groups from being recommended until quality can be assessed
- > limiting the number of Group invites a user can send
- > requiring mandatory post approval by administrators for Groups with low ‘integrity signals’
- > turning off commenting on particular posts when hate speech is detected
- > removing non-recommendable Groups that a user has joined as a factor for new groups recommended to that user.

4 Aug  **Lebanon** – Beirut explosion.

4 Aug ○○○ **Adobe, Microsoft** and *The New York Times* publish a white paper on the Content Authenticity Initiative. First announced in November 2019, the initiative aims to develop an industry standard for digital content attribution. As part of this work, Adobe is developing a system for creators and publishers to embed visual and audio-visual content with attribution data, creating an ‘attribution trail’ as changes are made to pieces of content.¹⁶⁵









¹⁶¹ ACMA, [Misinformation and news quality on digital platforms in Australia: A position paper to guide code development](#), June 2021

¹⁶² Leah Nysten, ‘[Google Announces steps to counter spread of hacked materials before election](#)’, *Politico*.

¹⁶³ Mike Schulman, ‘[Updates on our work to improve user privacy in digital advertising](#)’, Google Ads & Commerce Blog.

¹⁶⁴ Facebook, [Taking action to combat misinformation, polarization, and dangerous organizations](#).

¹⁶⁵ [The Content Authenticity Initiative: Setting the Standard for Digital Content Attribution](#) was co-authored by Adobe with BBC, CBC/Radio-Canada, Microsoft Corporation, The New York Times Company, Stanford Center for Blockchain Research, Truepic, University of California, Berkley and WITNESS.

- 4 Aug  **Facebook** pilots a new feature in WhatsApp that labels messages that have been forwarded at least 5 times with a magnifying glass icon. Users can click on this icon to launch a web search to help them verify the contents of the message.¹⁶⁶
- 5 Aug  **TikTok** announces a number of measures to address misinformation, disinformation and interference with the 2020 US presidential election.¹⁶⁷ These aim to:
- > prohibit manipulated content that may mislead users by distorting the truth of events in a manner that could cause harm (this was in addition to a pre-existing prohibition on content from disinformation campaigns)
 - > expand its fact-checking partnerships
 - > add an election misinformation option to the TikTok reporting feature to enable easy reporting of content or accounts for review
 - > introduce an election information centre to connect users to authoritative information
 - > work with the US Department of Homeland Security Countering Foreign Influence Task Force to help stop the threat and dangers of foreign influence on elections.
- 6 Aug  **Facebook** removes a coordinated inauthentic behaviour network that operates from multiple regions, including Australia, linked to the digital media outlet Truthmedia.¹⁶⁸
- 6 Aug  **US** – President Trump issues an Executive Order that would require TikTok to sell or spin off parts of its business to continue operating in the US. This order was revoked and replaced by a review of apps with ties to ‘jurisdiction of foreign adversaries’ initiated by President Biden in June 2021.
- 7 Aug  **Facebook** announces the winners of its call for ‘requests for proposals’ to study misinformation and polarisation.¹⁶⁹
- 8 Aug  **WhatsApp** releases new verification measures.¹⁷⁰
- 11 Aug  **Facebook** adds 2 new fact-checking labels: Altered and Missing Content, and updates its warning labels for Partly False and Missing Context ratings. Facebook maintains partnerships with more than 70 fact-checking organisations internationally. In Australia, Facebook partners with the Australian Associated Press and Agency France Presse, which are certified by the International Factchecking Network.¹⁷¹
- 13 Aug  **YouTube** makes changes to its policies to improve reliability of US election-related news,¹⁷² including to remove election-related content that violates its

¹⁶⁶ Manish Singh, ‘[WhatsApp pilots new feature to fight misinformation: Search the web](#)’, *TechCrunch*.

¹⁶⁷ Vanessa Pappas, General Manager, TikTok US, [Combating misinformation and election interference on TikTok](#).

¹⁶⁸ Facebook, ‘[July 2020 Coordinated Inauthentic Behaviour Report](#)’, *Facebook Newsroom*, 6 August 2020.

¹⁶⁹ Alex Leavitt, Kathryn Grant, ‘[Announcing the winners of Facebook’s request for proposals on misinformation and polarization](#)’, *Facebook Research Blog*.


¹⁷⁰ Facebook, [Taking action to combat misinformation, polarization, and dangerous organizations](#).


¹⁷¹ Facebook’s [submission to the Senate Select Committee on Foreign Interference through Social Media](#), p. 10; Facebook, [Response to the Australian disinformation and misinformation industry code](#), May 2021.

¹⁷² Leslie Miller, [An update on how YouTube supports elections](#), YouTube Official Blog.


policies (such as hacked information), removing content that encourages others to interfere with democratic processes, and labelling information about US electoral candidates.


Instagram introduces new authenticity measures, requiring accounts to confirm their information when Instagram identifies patterns of potentially inauthentic behaviour.


18 Aug  **US** – State of Emergency declared in California due to wildfires.

19 Aug  **Facebook** expands its Dangerous Organisations and Individuals policy to include measures to limit the spread of content on Facebook and Instagram from Groups that do not meet the definition of a 'dangerous organisation', such as QAnon and other US-based militia organisations.¹⁷³ Under the expanded policy, Facebook may take the following actions:


- > remove Facebook pages, Groups and Instagram accounts where discussions of potential violence are identified
- > limit recommendations for associated Facebook pages and Groups and Instagram accounts
- > rank content from these pages and Groups lower in the News Feed (implemented 16 September 2020)
- > reduce the visibility in the search function
- > disable the related hashtag function on Instagram
- > prohibit advertising, the selling of products or using Marketplace and Shop, on content related to these movements (on 29 September 2020 this was expanded to a prohibition against praising QAnon or militarised social movements)¹⁷⁴
- > prohibit identified Groups from using the service to fundraise.

22 Aug  **Australia** – Northern Territory general election.

20 Aug  **Twitter** expands its account labels to identify accounts for government officials, including heads of state, and state-affiliate media to the 5 permanent members of the UN Security Council.¹⁷⁵

31 Aug  **Facebook** announces research program on the impact of Facebook and Instagram on key political attitudes and behaviours during the US 2020 elections.

September 2020







1 Sep  **Microsoft** announces 2 AI technologies to assist with identifying deepfakes.¹⁷⁶ The first is the Microsoft Video Authenticator, which can assess the likelihood that an image or video has been manipulated. The second is a Microsoft Azure authenticity tool that allows content creators to add 'digital hashes and certificates' to a piece of content. Microsoft also announces its partnership with several media companies, including the BBC, CBC/Radio-Canada and *The New York Times*, to test the authenticity tool as part of its Project Origin initiative.

¹⁷³ Facebook, [An Update to How We Address Movements and Organizations Tied to Violence](#).

¹⁷⁴ Ibid.

¹⁷⁵ Twitter Support, [Expanding our work to identify state-affiliate accounts](#).

¹⁷⁶ Tom But and Eric Horvitz, [New Steps to Combat Disinformation](#).

- 3 Sep  **Facebook** announces it will label content which seeks to delegitimise the outcome of the US election (including from campaigns or candidates) or call premature victory.¹⁷⁷
- Facebook** introduces forwarding limits on **Messenger** to slow the spread of viral misinformation. Users can only forward messages to 5 people or groups at a time.¹⁷⁸
- 10 Sep  **EU – EC Staff Working Assessment of the EU Code of Practice on Disinformation.**
- 10 Sep  **Google** makes changes to its auto-complete policies in Search to remove false predictions about candidates in the US presidential election.¹⁷⁹
- 10 Sep  **Google** announces its Intelligence Desk, which is a global team of analysts set up to monitor news and events and assess how Google systems are performing against evolving news ‘narratives’, such as in relation to COVID-19. Google employs a number of systems to elevate authoritative sources of information, including surfacing fact-checked information and elevating original reporting on Google Search, providing information and knowledge panels that highlight credibility on particular issues, using machine learning to recognise authoritative sources as well as providing additional safeguards during breaking news or crisis events, such as the Breaking News shelf on YouTube, which appears on the YouTube home page when a significant news event occurs and features authoritative sources of information. Google states that it continues to ‘improve its systems across Google and YouTube so that [it can] detect breaking news contexts (and crisis situations) and optimise for elevating authoritative sources’ and reduce the spread of misinformation.¹⁸⁰
- 16 Sep  **Facebook** expands its ‘downranking’ to pages and Groups that have been restricted but not removed.
- 17 Sep  **Facebook** implements further changes to how it manages Groups to further reduce the spread of misinformation.¹⁸¹ The changes include:
- > preventing administrators/moderators of Groups that have contravened Facebook’s policies from creating any new Groups for a period of time
 - > requiring posts from Group members who have contravened Facebook’s policies be approved by the Group administrator for 30 days
 - > archiving Groups without active administrators/moderators
 - > removing health-related Groups from Group recommendations
 - > removing Groups that share misinformation in contravention of Facebook’s Community Guidelines
 - > downgrading rankings of Groups that share false information (identified by fact-checkers) in recommendations and News Feeds
 - > labelling content that has been reviewed by fact-checkers.


¹⁷⁷ Facebook, [New steps to protect the US elections](#).

¹⁷⁸ Facebook, [Introducing a Forwarding Limit on Messenger](#), 3 September 2020.


¹⁷⁹ Google, [Our latest investments in information quality in Search and News](#).

¹⁸⁰ Google, [Australian code of practice on disinformation and misinformation: Google Initial Report, May 2021](#).

¹⁸¹ Tom Alison, [Our latest steps to keep Facebook Groups safe](#).


17 Sep  **Twitter** increases account security protections for a designated group of high-profile US election-related accounts.¹⁸² The group included accounts relating to the US Executive Branch and Congress, US Governors and Secretaries of State, Presidential campaigns, political parties, and candidates with Twitter Election Labels running for the US House, Senate, or for US state Governor, and major US news outlets and political journalists. Other accounts could also make use of the new measures, which include:

- > requiring strong passwords
- > password reset protection enabled by default
- > 2-factor authentication strongly encouraged
- > internal security measures increased, such as more sophisticated alerts of suspicious activity.

17 Sep  **Twitter** updates its Civic Integrity Policy to specifically address false and misleading information 'intended to undermine public confidence in an election or other civic process'.¹⁸³ Under the policy, Twitter may remove or label tweets that make false or misleading claims:


- > about laws underpinning civic processes
- > which may undermine faith civic processes
- > about the outcome of a civic process with the intention of interfering with the that outcome.


Tweets labelled under this policy will have their visibility reduced.

18 Sep  **Twitter** releases its Coordinated Harmful Activity policy¹⁸⁴ that sets out the framework by which it assesses whether a group, movement or campaign is engaged in Coordinated Harmful Activity. The framework defines the relevant categories of harm, physical, psychological and informational, consistent with other policies in the Twitter Rules. Coordination is defined as either:

- > technical coordination, which 'refers to the use of specific detectable techniques of platform manipulation to engage in the artificial inflation or propagation of a message or narrative' (all forms are prohibited under the Twitter rules)
- > social coordination, which 'refers to on or off Twitter coordination among a group of people to amplify or propagate a specific message' (some forms constitute a violation of the Twitter rules).

Under the framework, if Twitter identifies Coordinated Harmful Activity, it may take a number of enforcement actions on tweets and accounts, including to limit their visibility and suspend accounts primarily used for Coordinated Harmful Activity. Under the Twitter rules more generally, tweets may be removed that violate its policies and accounts may be suspended for severe or repeated violations of the Twitter rules.

22 Sep  **TikTok** launches an Asia-Pacific Safety Advisory Council, made up of independent experts to advise on TikTok's content moderation policies as well as broader matters relating to trust and safety at TikTok.¹⁸⁵

23 Sep  **WHO** issues a joint statement on managing the COVID-19 'infodemic'.

¹⁸² Twitter Safety, [Improved account security during the 2020 US election](#).





¹⁸³ Twitter Safety, [Expanding our policies to further protect the civic conversation](#).

¹⁸⁴ Twitter policy, [Coordinated harmful activity](#).

¹⁸⁵ Arjun Narayan Bettadapur, [Introducing the TikTok Asia Pacific Safety Advisory Council](#).

- 25 Sep  **TikTok** appears before Australian Senate Select Committee on Foreign Interference through Social Media.
- 25 Sep  **Google** announces that it will block election advertisements globally after 3 November 2020 until 7 days after the 2020 US Presidential election day.¹⁸⁶ This included advertisements that were explicitly election related (an advertisement is considered election related if it mentioned a current state of federal officeholder or candidate, political party, or ballot measure), any other advertisements that reference federal or state elections, or advertisements that run based on targeting election-related search queries.
- 29 Sep  **Facebook** prohibits advertising that support militarised social movements and QAnon.
- 30 Sep  **US** – First US Presidential debate.
- 30 Sep  **Facebook** implements new measures to limit the activities of QAnon. When users search for certain hashtags related to child safety, they will be redirected to credible information. QAnon content identified as false by third-party fact-checkers will also be limited in News Feed, filtered from Explore and hashtags on Instagram, and labelled with additional contextual material.¹⁸⁷

October 2020

- 1 Oct  **TikTok** launches fact-checking in Australia with the global news agency AFP.¹⁸⁸ TikTok has fact-checking partners across 8 markets that assist TikTok to remove misinformation from its platform. Users are also able to report information that they believe may violate its misleading content policies, using the ‘misleading information’ category.
- 2 Oct  **US** – President Trump tests positive for COVID-19.
- 6 Oct  **Facebook** makes further updates to its Dangerous Individuals and Organisations policy on QAnon, to remove any Pages, Groups, and Instagram accounts representing QAnon even if they do not contain violent content. Facebook states that it is proactively detecting content that may contravene this policy, rather than relying on user reports. Facebook notes that the policy change was linked to evidence that QAnon content was spreading other kinds of misinformation (other than content promoting violence), leading to real world harm, such as claims that the US west coast wildfires were started by certain groups.¹⁸⁹
- 7 Oct  **Microsoft** announces new initiatives to support journalism and local newsrooms.¹⁹⁰ Microsoft has commenced 4 pilot programs in the US in which it will provide direct funding, improve available technology, help build capacity around technological transformation, expand news distribution and coordinate knowledge-sharing on successful approaches to revenue and funding.

¹⁸⁶ Sara Fischer, [Scoop: Google to block election ads after Election Day](#), Axios.








¹⁸⁷ Facebook, [An Update to How We Address Movements and Organizations Tied to Violence](#).

¹⁸⁸ Arjun Narayan Bettadapur, [TikTok partners with fact-checking experts to combat misinformation](#).

¹⁸⁹ Facebook, [An Update to How We Address Movements and Organizations Tied to Violence](#).

¹⁹⁰ Mary Snapp, [New steps to preserve and protect journalism and local newsrooms](#).

Microsoft is also piloting the Protecting Journalists Pro Bono Program to provide legal support to news organisations.















- 7 Oct  **Facebook** announces measures to prevent misinformation in relation to the US presidential election results. These include notifications at the top of Facebook and Instagram as well as on the posts of candidates, and directing users to its Voter Information Center. Facebook states that, in the event that a candidate or party called premature victory before the election had been called by a major media outlet, it intends to include additional information about the counting process in warning labels. If the election result is contested, Facebook states that it would include information about the declared winner in a notification at the top of Facebook and Instagram.
- 8 Oct  **Twitter** discloses 5 networks of accounts to its archive of potential information operations (accounts of these networks suspended for violations of Twitter's policies).
- 9 Oct  **Twitter** announced updates to its Civic Integrity Policy to mitigate possible attempts to interfere with the outcome of the US Presidential election. The updates prohibit all users from making claims about who won the election before the outcome is called by an authoritative source. Under the policy, an authoritative source refers to an announcement from US state election officials, or a public projection from at least 2 national news outlets that make independent election calls. The policy also prohibits tweets intended to incite interference with the results.¹⁹¹ Twitter also announces a series of other measures to support its expanded policy, with the intention of increasing the context of tweets and encouraging users to consider more thoughtfully consider retweeting information. For example, users will be prompted to view credible information before being able to retweet a tweet with a misleading information label.¹⁹²
- 13 Oct  **Facebook** launches a global policy prohibiting advertising that discourages vaccination.¹⁹³ Under the policy, advertising that advocates a particular position on the government's policy on vaccines would not necessarily contravene Facebook's Community Standards, provided they include a 'Paid for by' label.
- 14 Oct  **US** – Hunter Biden's hacked emails published by *The New York Post*.
- 14 Oct  **Facebook** and **Twitter** limit spread of a *New York Post* article on Hunter Biden.
- 14 Oct  **Google** updates its COVID-19 medical misinformation policy for YouTube to address vaccines.¹⁹⁴

¹⁹¹ Blog post by Vijaya Gadde and Kayvon Beykpour, [Additional steps we're taking ahead of the 2020 US Election](#).

¹⁹² Ibid.

¹⁹³ Kany-Xi Jon, [Supporting Public Health Experts' Vaccine Efforts – About Facebook](#).

¹⁹⁴ Google, *Australian code of practice on disinformation and misinformation: Google Initial Report, May 2021*.

16 Oct		Twitter makes changes to its Hacked Materials Policy to no longer remove hacked content unless it is directly shared by hackers or those acting in concert with them. ¹⁹⁵ Under the policy, Twitter will label tweets to provide context instead of blocking links from being shared on Twitter.
17 Oct		New Zealand – General election.
		Australia – ACT general election.
		Twitter introduces new feature to its platform, prompting users before they retweet or quote content that is labelled as containing misleading information.
18 Oct		TikTok removes content and accounts that promote QAnon for violating its Community Guidelines.
20 Oct		Facebook temporarily suspends civic and political Groups being recommended in the US. ¹⁹⁶
21 Oct		Facebook redirects searches on QAnon (and related issues) to credible resources from the Global Network on Extremism and Technology (the academic research network of the Global Internet Forum to Counter Terrorism). ¹⁹⁷
22 Oct		Facebook's Oversight Board officially begins to hear cases with the intention of prioritising cases with the greatest potential for impact on users globally. ¹⁹⁸
23 Oct		TikTok updates its notification settings globally so that creators of content that violates its policies will be notified of the policy they have contravened and provide them with the ability to appeal the decision. ¹⁹⁹
26 Oct		Facebook stops accepting new political advertising the week before the US presidential election.
27 Oct		US – Section 230 of the <i>Communications Decency Act 1996</i> (US) hearing: Facebook, Google and Twitter appear before the US Senate Commerce committee
30 Oct		Facebook suspends political group recommendations in the US.
31 Oct		Australia – Queensland state election.
November 2020		
		Europe – COVID-19 second wave

¹⁹⁵ [Tweet](#) by Vijaya Gadde, Twitter Legal, Policy and Trust & Safety Lead.

¹⁹⁶ Ryan Mac and Craig Silverman, '[Facebook Quietly Suspended Political Group Recommendations Ahead Of The US Presidential Election](#)', *BuzzFeed*.

¹⁹⁷ Facebook, [An Update to How We Address Movements and Organizations Tied to Violence](#).

¹⁹⁸ Brent Harris, [Oversight Board to Start Hearing Cases – About Facebook](#).

¹⁹⁹ TikTok, [Adding clarity to content removals](#).



Indonesia – Passes Ministerial Regulation No.5 (MR%) which requires Private Electronic System Operators to register with the government and make content accessible for monitoring. The regulation also enables the government to order the removal of prohibited content, including content which ‘promotes social anxiety and disrupts public order’.²⁰⁰

3 Nov



Google blocks election advertising in the US.

Facebook temporarily stops all social issue, electoral or political advertising in the US.

4 Nov



US – Presidential election day. Allegations of ‘election fraud’ (including from President Trump and QAnon) are widespread.

4 Nov



Google displays election results from The Associated Press when users search for election results.²⁰¹

5 Nov



Facebook removes the ‘Stop the Steal’ Group under its Coordinating Harm policy.

5 Nov



Facebook is reported to (temporarily) modify its News Feed to prioritise mainstream media content, in the days immediately following the US Presidential election.²⁰²

8 Nov



UK – UK government agree on a set of measures with Google, Facebook, and Twitter to address COVID-19 vaccine disinformation and misinformation and to promote reliable information.

16 Nov



Australia – COVID-19 outbreak in Adelaide and short-term lockdown.

18 Nov



YouTube includes links to authoritative information on COVID-19 vaccines to its COVID-19 information panels. Users who click on the links are taken to authoritative third-party sources, like the Centers for Disease Control and Prevention or the WHO. Information panels are a feature used by Google to provide additional information and context to matters it identifies as prone to misinformation.²⁰³

25 Nov



Google announces that its political advertising reporting functionality and advertising library will become available in Australia. Election advertisements in Australia are advertisements that feature a political party, current elected officeholder or candidate for the House of Representatives or Senate. Election advertisements do not include advertisements for products or services, including promotional political merchandise, such as t-shirts or advertisements run by news organisations to promote their coverage of political parties, candidates, or current elected officeholders. The new function allows users to see how much money parties and other groups spend on targeted election advertisements.


²⁰⁰ Eduard Lazarus, ‘The authoritarian threat of Indonesia’s latest internet bill’, [The Interpreter](#), 7 June 2021.

²⁰¹ Amanda Storey, [Following the 2020 US Election with Google](#).


²⁰² Kevin Roose, Mike Isaac and Sheera Frenkel, [‘Facebook Struggles to Balance Civility and Growth’](#).


²⁰³ YouTube Help, [Information panel giving topical context](#).

Google also introduces new policies for election advertising in Australia, requiring that they be labelled with a 'paid for by' disclosure. Advertisers must also go through a verification process. Political advertising is also subject to Google's other advertising policies. Google states that targeting election advertising is permitted based on general geographic post code, age, gender, and context. Targeting is not permitted based on political affiliation.²⁰⁴


28 Nov  **TikTok** announces it will partner with fact-checkers for the US presidential election results, to help reduce discoverability of content that prematurely claims victory in a race before results are confirmed by the Associated Press.²⁰⁵ TikTok also displays a banner on content with unverifiable claims about premature declarations of victory.


December 2020

2 Dec  **Twitter** expands its Hateful Conduct Policy to address content that dehumanises people on the basis of race, ethnicity, or national origin.²⁰⁶ Any content that violates this policy is removed through proactive detection and automatic moderation.

3 Dec  **Facebook** announces that it will also apply existing policies to remove misinformation that can lead to imminent, physical harm, or false claims about COVID-19 vaccines. Facebook advised that, in enforcing this policy, it would regularly update the types of false claims that it would remove under its policies.²⁰⁷

3 Dec  **EU – EC** presents its Democracy Action Plan.

9 Dec  **YouTube** starts applying its Presidential Election Integrity Policy, under which content that alleges the outcome was the result of widespread fraud will be removed or errors will be removed (this policy does not relate to elections outside the US).²⁰⁸ For the 2020 US presidential election, this policy applies to content uploaded on or after the 9 December 2020.

10 Dec  **Google** launches a new feature in the UK to provide information on and surface a list of authorised vaccines in a user's location.²⁰⁹ This feature will be rolled out as other jurisdictions begin offering the vaccine.

○ ○ ○ **Google** provides a \$15 million 'Ad Grant' to the WHO to provide public services announcements about the pandemic.²¹⁰ Google has also provided \$4.8 million in Ad Grants to the Australian Federal Government and Department of Health and \$48 million to Australian not-for-profits,²¹¹ and \$250 million in Ad Grants to more than 100 government agencies throughout

²⁰⁴ Google, *Australian code of practice on disinformation and misinformation: Google Initial Report, May 2021*, p. 26.

²⁰⁵ Eric Han, Head of Safety, TikTok US, [Supporting our community on Election Day and beyond](#).

²⁰⁶ Twitter Safety blog, [Updating our rules against hateful conduct](#).

²⁰⁷ Kang-Xing Jin, [Keeping People Safe and Informed About the Coronavirus – About Facebook](#).


²⁰⁸ [Spam, deceptive practices, & scams policies – YouTube Help \(google.com\)](#).


²⁰⁹ Google, [How you'll find accurate and timely information on COVID-19 vaccines](#).


²¹⁰ Ibid.


²¹¹ Google, *Australian code of practice on disinformation and misinformation: Google Initial Report, May 2021*.


2020. Google also announces it will provide \$1.5 million for the creation of a COVID-19 Vaccine Media Hub by the Australian Science Media Centre.²¹²


14 Dec  **US** – Joseph R. Biden confirmed as President-elect by the Electoral College.

15 Dec  **UK** – Government publishes *Online Harms White Paper response*.
EU – EU Commission publishes the *Digital Services Act and the Digital Markets Act*.


15 Dec  **TikTok** updates its in-app coronavirus resource hub with commonly asked questions and answers about COVID-19 vaccines from public health experts.²¹³ These changes are made as part of broader updates to TikTok's policies to support well-being.
Facebook introduces notifications to alert people if they had interacted with a piece of COVID-19-related content that had been removed under its misinformation policies, explaining why it was false and linking to credible information.²¹⁴


16 Dec  **Twitter** expands policies to address misinformation about COVID-19 vaccines.²¹⁵ Twitter requires people to remove tweets that advance harmful, false or misleading narratives about vaccines, such as false or misleading claims that vaccines are used for harmful purposes, have adverse effects or are unnecessary. From early 2021, Twitter may put warning labels on tweets that spread potentially misleading information about vaccines.²¹⁶

18 Dec  **Australia** – COVID-19 cluster and local lockdown in Sydney's northern beaches. Restrictions are lifted 10 January 2021.

23 Dec  **Twitter** introduces an emergency search prompt, in collaboration with the Australian Red Cross. Australian users who search for keywords, such as bushfire, flooding and cyclone, are directed to the Australian Red Cross Twitter account and resources.²¹⁷

January 2021

6 Jan  **US** – Storming of US Capitol building, following the 'Save America Rally' in Washington DC.

6 Jan  **Facebook** temporarily restricts several Group features, including:

- > requiring administrators to review and approve posts before they are visible in certain Groups
- > automatically disabling comments where a high rate of hate speech or content that incites violence is detected in a Group
- > using artificial intelligence to downrank content that is likely to contravene Facebook policies.

²¹² Google, [How you'll find accurate and timely information on COVID-19 vaccines](#).


²¹³ Cormac Keenan, Head of Trust & Safety, TikTok, [Refreshing our policies to support community well-being](#).

²¹⁴ Guy Rosen, [An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19 – About Facebook](#).


²¹⁵ Twitter Safety, [COVID-19: Our approach to misleading vaccine information](#).


²¹⁶ Twitter Safety, [Updates to our work on COVID-19 vaccine misinformation](#).

²¹⁷ Kara Hinesley, [Twitter launches a search prompt with the Australian Red Cross](#).


- 


Facebook suspends President Trump's account for 24 hours.
- 7 Jan




YouTube removes President Trump's address on the storming of the Capitol, including his description of the perpetrators as 'special' under its Presidential Election Integrity Policy.
- 


Facebook suspends President Trump's Facebook and Instagram accounts.
- 8 Jan




Twitter permanently bans President Trump from its platform.
- 

TikTok removes President Trump's speeches, which reiterate claims of a fraudulent election.
- 


Google removes the Parler app from the Google Play Store.
Apple follows suit the following day, removing Parler from the App Store on 9 January.
- 11 Jan




Facebook removes content including 'Stop the Steal' under its Coordinating Harm policy.²¹⁸
- 12 Jan




Twitter updates its Civic Integrity Policy to increase enforcement measures, including that repeated violations can result in permanent suspension of an account.²¹⁹
- 16 Jan




Facebook temporarily bans weapon accessories and protective equipment advertising until 22 January 2021.²²⁰
- 20 Jan




US – Inauguration of Joe Biden as 46th US President.
- 21 Jan



Facebook refers its decision to indefinitely suspend President Trump from Facebook and Instagram to its Oversight Board for independent consideration due to its significance.²²¹
- 25 Jan



Twitter introduces 'Birdwatch' pilot program, which allows users to tag and comment on information they think might be misleading. Twitter aims to make the program available globally.²²²
- 27 Jan



Facebook permanently bans civic and political Groups from being recommended and rolls this policy out globally.

February 2021

- 

Google expands its collaboration with Defending Digital Campaigns.

²¹⁸ Facebook, [Our Preparations Ahead of Inauguration Day – About Facebook](#).

²¹⁹ Twitter Safety, [An update following the riots in Washington, DC](#).

²²⁰ Guy Rosen, [Our Preparations Ahead of Inauguration Day – About Facebook](#).

²²¹ Nick Clegg, [Referring Former President Trump's Suspension From Facebook to the Oversight Board – About Facebook](#).

²²² Twitter, [Introducing Birdwatch, a community-based approach to misinformation](#), 25 January 2021.

○ ○ ○ **Adobe, Microsoft, Arm, BBC, Intel, and Truepic** form the Coalition for Content Provenance and Authenticity (C2PA). The purpose of C2PA is to create open technical standards for certifying the origin and evolution history of digital media content as a means of addressing disinformation and misinformation. C2PA brings together the Content Authenticity Initiative and Project Origin (originally focused on online news distribution), which will continue to engage with their respective sectors, under unified standards.



Facebook publishes a list of specific claims it considers to be misinformation that could cause imminent, physical harm.²²³ It also expands its Misinformation and Harm policy to include false and misleading claims about vaccines more generally, such as that vaccines cause autism or SIDS, or other alarmist content related to vaccines.

○ ○ ○ **TikTok** launches a 'Know the Facts' campaign in Australia.²²⁴

8 Feb



Facebook, following consultation with the WHO and other health organisations, expands the scope of its policy to remove false content on COVID-19 vaccines on Facebook and Instagram,²²⁵ such that Facebook and Instagram Groups, pages and accounts that repeatedly share false claims (as determined by third-party fact-checkers) could be removed.

8 Feb



Facebook announces that it will grant \$120 million in advertising credits to health agencies (including in Australian federal and state agencies), NGOs and UN agencies to provide people with information on the COVID-19 vaccine and preventive health. Facebook also makes modifications to its search function to return higher instances of validated vaccine information and reduce identified misinformation.

16 Feb



Facebook temporarily bars Australians from finding or sharing news on the platform, in response to the Australian Government's News Media and Digital Platforms Bargaining Code. The ban is lifted a week later on 23 February.

17 Feb



Facebook and **Instagram** permanently ban celebrity chef Pete Evans for repeatedly sharing misinformation about COVID-19 and vaccines.

17 Feb



Twitter expands its account labels to identify accounts for government officials, including heads of state, and state-affiliate media for G7 nations.²²⁶ Twitter states that it intends to expand this policy.²²⁷

22 Feb



Australia – DIGI launches the *Australian Code of Practice on Disinformation and Misinformation* and announces the 6 initial code signatories.



Australia – COVID-19 vaccine rollout commences.



²²³ Facebook, *Facebook response to the Australian disinformation and misinformation industry code*, p. 26.

²²⁴ TikTok, *Australian Code of Practice on Disinformation and Misinformation: Initial Report, May 2021*.








²²⁵ Margaret Harding McGill and Sara Fischer, '[Facebook says it will crack down on COVID vaccine misinformation](#)', *Axios*.

²²⁶ Twitter Support, [Expanding our work to identify state-affiliated accounts](#).

²²⁷ Twitter, *Twitter: Australian Code of Practice on Disinformation and Misinformation Initial Report*, p. 18.

- 23 Feb  **Twitter** discloses 5 networks of accounts to its archive of potential information operations (accounts of these networks suspended for violations of Twitter’s policies).
- 25 Feb  **India** – Introduces the *Information Technology (Guidelines for Intermediaries and Digital Media Ethics Code) Rules* requiring social media and digital streaming companies to remove content deemed unlawful.

March 2021

- 1 Mar  **TikTok** introduces a new feature, allowing users to decide which comments appear next to a post. This feature is in addition to existing controls that enable users to filter spam and offensive comments, and specific keywords. The new feature is also complemented by a prompt to TikTok’s Community Guidelines, and suggests reconsidering posting a comment, when TikTok detects that a comment may be ‘inappropriate or unkind’.²²⁸
- 1 Mar  **Twitter** updates its COVID-19 policy to include an enforcement strike system, designed to help to educate the public on its policies, and further reduce the spread of potentially harmful and misleading information on Twitter, particularly for repeated moderate and high-severity violations of its rules.²²⁹
- 4 Mar  **Facebook** implements controls enabling Facebook and Instagram users to limit electoral, political and social advertising in 90 countries.²³⁰ In Australia, the feature is available for political- and election-related advertising only.²³¹
- 13 Mar  **Australia** – Western Australia state election.
- 15/16 Mar  **Facebook** expands on its 15 December 2020 notifications by also applying information labels to content discussing COVID-19 vaccines on Facebook and Instagram.²³² The labels include credible information from the WHO on the safety and efficacy of COVID-19 vaccines. Facebook also launches a COVID-19 Information Center on Instagram on 16 March.²³³
- 17 Mar  **Facebook** implements further restrictions to Groups, including that Groups that have contravened the Community Guidelines will appear lower down in recommendations (Groups that repeatedly contravene Facebook’s Community Guidelines are removed).²³⁴ Facebook will notify people when they are joining a Group that has violations and limit invite notifications for these Groups. Where Facebook identifies ‘severe harm’, it will remove Groups without implementing gradual enforcement.
- 18 Mar  **Australia** – NSW floods start. Floods subside by the end of March 2021.

²²⁸ Tara Wadhwa, ‘New Tools to promote kindness on TikTok’, [TikTok Safety](#), 11 March 2021.

²²⁹ Twitter Safety, [Updates to our work on COVID-19 vaccine misinformation](#), 1 March 2021.




²³⁰ Naomi Cleit, [Launching the largest voting information effort in US history](#).

²³¹ [Availability for ads about social issues, elections or politics | Facebook Business Help Centre](#).






²³² Facebook, ‘Mark Zuckerberg announces Facebook’s plans to help get people vaccinated against COVID-19’, [Facebook newsroom](#), 15 March 2020.

²³³ Instagram, ‘Helping people stay safe and informed about COVID-19 vaccines’, [Instagram Announcements](#), 16 March 2021.

²³⁴ Tom Alison, [Changes to keep Facebook Groups Safe](#).

18 Mar		Apple News features special curated section on the flood emergency. ²³⁵
24 Mar		Instagram introduces a range of updates to improve COVID-19 information accuracy, including only recommending COVID-19 related accounts if they belong to a credible health organisation.
31 Mar		Facebook introduces a new feature to give users greater ability to control comments on their posts. ²³⁶ The change is available to all users but will enable news media organisations to more effectively moderate comments for potentially defamatory content.

April 2021

Early Apr		India – Beginning of COVID-19 second wave.
		Twitter introduces a timeline prompt to link Australian users with the Australian Government Department of Health's COVID-19 landing page providing information on vaccines and vaccination. ²³⁷
8 Apr		US – Facebook, Google and Twitter appear before the House Committee on Energy and Commerce on the role and responsibility of digital platforms in promoting extremism and misinformation following the 2020 US election.
12 Apr		Australia – Australian Government updates official health advice regarding the AstraZeneca COVID-19 vaccine (Pfizer preferred for under 50s).
13 Apr		Facebook -funded report on media literacy in Australia launched by the Western Sydney University, Queensland University of Technology and the University of Canberra. ²³⁸
14 Apr		Facebook's Oversight Board starts accepting appeals from Facebook and Instagram users that want content removed from these services. Previously, users were only able to appeal to have content restored. ²³⁹
19 Apr		Facebook announces proactive measures to limit real-world harm arising from the Derek Chauvin verdict on the murder of George Floyd. Facebook indicates that, as temporary emergency measures, it may limit the spread of content that its systems predict will contravene its policies, and remove events organised in temporary, high-risk locations that contain calls to bring arms. ²⁴⁰
26 Apr		Facebook permanently removes Craig Kelly MP's Facebook and Instagram accounts for repeated violations of its COVID-19 misinformation policy.

²³⁵ Apple, *Apple News: Australian Code of Practice on Disinformation and Misinformation Initial report*.

²³⁶ Josh Taylor, [Facebook now lets users and pages turn off comments on their posts](#), *The Guardian*, 31 March 2021.

²³⁷ Twitter, *Twitter: Australian Code of Practice on Disinformation and Misinformation Initial Report*, p. 8.


²³⁸ Facebook, *Facebook response to the Australian disinformation and misinformation industry code*.

²³⁹ Oversight Board, [The Oversight Board is accepting user appeals to remove content from Facebook and Instagram](#), April 2021.

²⁴⁰ Monica Bickert, [Preparing for a Verdict in the Trial of Derek Chauvin](#), Facebook Newsroom, 12 April 2021.


A Facebook spokesperson noted: 'We don't allow anyone, including elected officials, to share misinformation about COVID-19 that could lead to imminent physical harm or COVID-19 vaccines that have been debunked by public health experts. We have clear policies against this type of content and have removed Mr Kelly's Facebook Page for repeated violations of this policy'.²⁴¹

27 Apr ○○○ **TikTok** announces it will open a European Transparency and Accountability Centre, similar to its US Transparency Center. Independent experts will be able to observe how TikTok implements its content moderation policies.²⁴²

29 Apr  **Facebook** temporarily hides posts containing #ResignModi due to some posts violating Community Standards. Facebook notes this was a mistake and restores posts with this hashtag after several hours.²⁴³

May 2021

1 May  **Australia** – Tasmanian state election.

5 May  **Facebook's** Oversight Board upholds Facebook's decision to suspend President Trump from Facebook and Instagram but determines that indefinite suspension is inconsistent with Facebook's normal penalties. The Board insists that Facebook review the matter within 6 months to apply a 'proportionate response that is consistent with the rules that are applies to other users of its platform'.²⁴⁴

Facebook announces its response to the Oversight Board finding on 4 June, noting that it will suspend former President Trump's account for 2 years, ending 7 January 2023. The suspension is the highest penalty under a new enforcement protocol developed by Facebook in response to the Oversight Board's findings. Facebook will also provide more information in its Transparency Center about when it applies its 'newsworthiness allowance' and allows content to stay on the platform on public interest grounds, where it might otherwise contravene Facebook's policies.²⁴⁵

12 May ○○○ **UK** – UK Government publishes its draft Online Safety Bill. This includes a proposal for independent regulator Ofcom to establish an advisory committee on disinformation and misinformation and publish periodic reports on this issue.

22 May ○○○ **Australia** – DIGI publishes the code commitments and transparency reports of all 8 signatories to the code. **Facebook** signals its intent to expand its fact-checking capability in Australia in the second half of 2021.²⁴⁶

²⁴¹ Jade Macmillan and Brett Worthington, [Facebook removes Craig Kelly's page, says former Liberal MP breached misinformation policies](#), ABC News, 26 April 2021.


²⁴² Natasha Lomas, [TikTok to open a 'Transparency' Center in Europe to take content and security questions](#), TechCrunch, 27 April 2021.

²⁴³ Kari Paul, [Facebook blocked hashtag calling for Narendra Modi to resign over pandemic](#), The Guardian, 29 April 2021.

²⁴⁴ Oversight Board, [Case decision 2021-001-FB-FBR](#), Board Decisions, 5 May 2021.

²⁴⁵ Nick Clegg, [In response to oversight board, Trump suspended for two years; will only be reinstated if conditions permit](#), Facebook Newsroom, 4 June 2021.

²⁴⁶ Josh Machin, [Facebook response to the Australian disinformation and misinformation industry code](#), Facebook Australia, 21 May 2021.

26 May  **WhatsApp** brings a lawsuit against the Indian Government over new laws that give the Indian Government powers to monitor online activity.²⁴⁷




EU – European Commission publishes its *Guidance on Strengthening the Code of Practice on Disinformation*, recommending an expansion of scope.



Facebook places further limits on users who repeatedly share false content that has been fact-checked, by reducing the distribution of all their posts in News Feed, rather than just reducing reach of individual posts.²⁴⁸ Facebook also lifts ban on claims that COVID-19 is man-made.²⁴⁹

June 2021

4 Jun  **Nigeria** indefinitely suspends Twitter following its deletion of a tweet by President Muhammadu Buhari and announces plans to prosecute any Nigerians defying the government's ban.²⁵⁰

²⁴⁷ Hannah Ellis-Petersen, [WhatsApp sues Indian government over 'mass surveillance' internet laws](#), *The Guardian*, 26 May 2021.

²⁴⁸ Facebook, [Taking Action Against People Who Repeatedly Share Misinformation](#), Facebook Newsroom, 26 May 2021.

²⁴⁹ Alex Hern, [Facebook lifts ban on posts claiming Covid-19 was man-made](#), *The Guardian*, 28 May 2021.

²⁵⁰ Associated Press, [Nigeria's 40 million Twitter users banned from site as government-enforced suspension takes effect](#), *ABC News*, 7 June 2021.

Appendix D: International regulatory approaches

Several jurisdictions have introduced or are considering anti-disinformation or misinformation initiatives relating to digital platforms.

To inform the development of this report, the ACMA has monitored a range of international regulatory developments and consulted with international regulators such as Ofcom, the Broadcasting Authority of Ireland, and the Canadian Radio-television and Telecommunications Commission, to share common experiences and learn from their approaches.

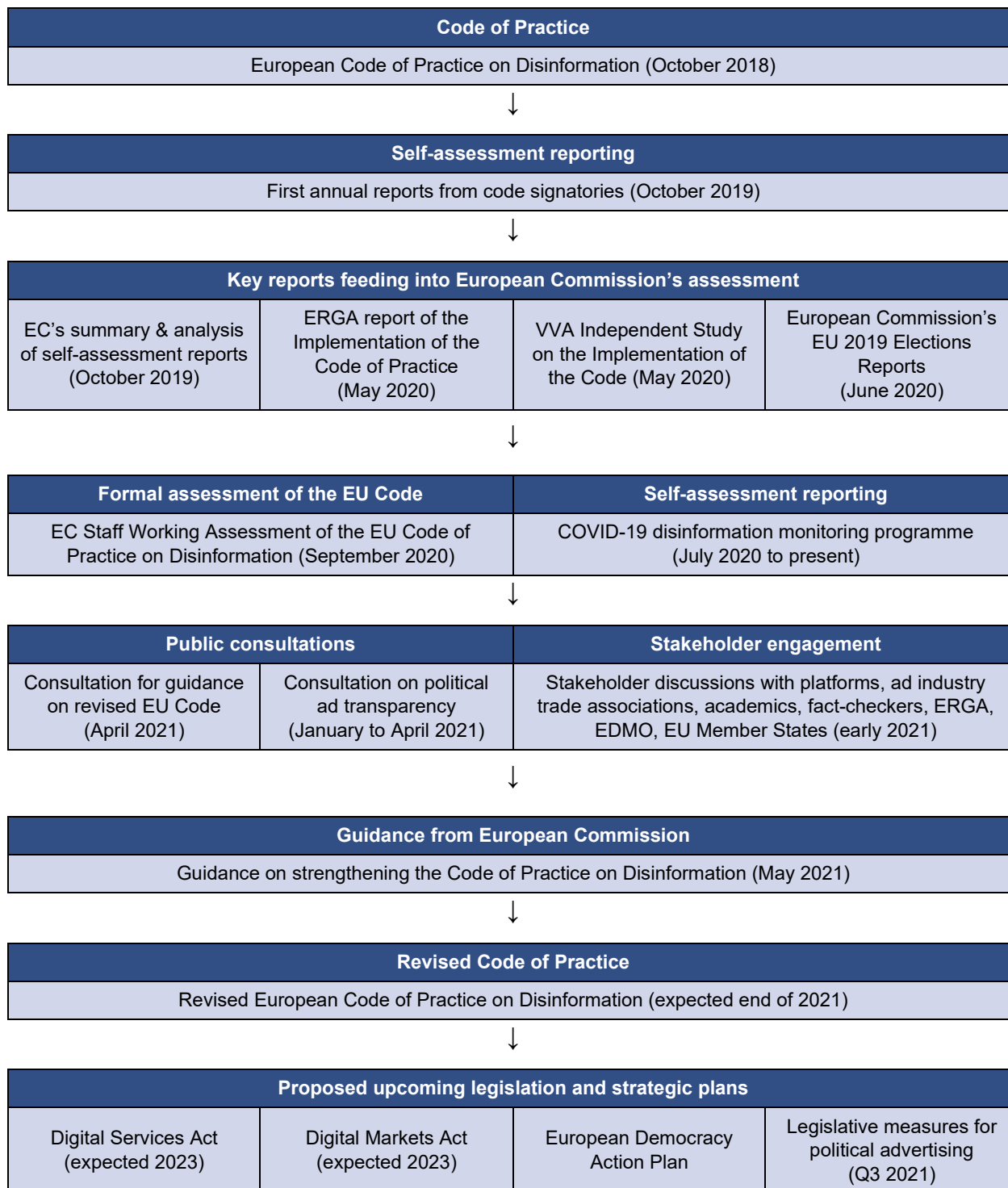
Europe

The Australian Government's response to the DPI asked digital platforms to draw on learnings from the EU Code to inform the development of the Australian code(s).

The EU Code is a voluntary, outcomes-based code that sets out a list of high-level commitments and principles that signatories agree to follow to protect users from disinformation. The code was introduced October 2018 as a pioneering self-regulatory code to address disinformation. Current signatories include Facebook, Google, Microsoft, Twitter, Mozilla, TikTok, and several advertising industry groups.

Since its inception, there have been significant developments to evaluate and redress the EU Code. An overview of these developments is listed below and a timeline is provided at Figure 21.

Figure 21: Timeline of key EU Code developments



European Commission's assessment of the EU Code

In September 2020, the European Commission published its formal assessment of the EU Code for its initial 12-month period from October 2018 to October 2019.²⁵¹ The assessment report is a 'staff working document' and does not contain any formal recommendations for the European Commission or code signatories.

The assessment found that the code is a valuable instrument and provides a framework for a structured dialogue between relevant stakeholders to ensure greater transparency of platforms' policies against disinformation in the EU.

Some of the key successes of the EU Code identified by the European Commission include:

- > greater transparency of political advertising, including clearly labelling political ads, identifying candidates, political parties, and the creation of ad libraries to contain all political ads served
- > platforms' removal of fake accounts, malicious bots, and spam content
- > greater investment in technology to give prominence to trustworthy information sources and make it easier for users to critically assess and find diverse perspectives about topics of public interest through new tools
- > new services and features for users to flag potential instances of disinformation and be warned about content that has been fact-checked and rated as false or misleading
- > new policies and tools that provide researchers and the fact-checking community with better access to platform data.

Weaknesses of the code were mostly attributed to limited participation, lack of independent oversight, and the need for further data to be provided from platforms to the research community and relevant regulatory authorities. Key shortcomings include:

- > lack of collaboration of stakeholders, including platforms, fact-checkers and advertisers
- > no data showing the effectiveness of news tools to increase user engagement with trustworthy information sources or enhance critical thinking
- > no user-friendly and uniform procedure available on all platforms for users to flag possible disinformation cases and be adequately informed about the outcome of their actions
- > inadequate approaches to raise users' awareness of content labelled as false or misleading, including when a user has previously come across this content
- > lack of information on the integration of trustworthiness indicators in platforms' algorithms.

The European Commission also identified some emerging areas in the disinformation environment that remain unaddressed in the code, including:

- > manipulative online behaviours, such as fake engagement techniques aimed at inflating the perceived popularity of certain content
- > micro-targeting of political advertising
- > KPIs and data for monitoring the code to address the lack of common reporting structures and allow for cross-platform comparisons.

²⁵¹ European Commission, [Commission Staff Working Document: Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement](#), September 2020.

COVID-19 disinformation monitoring program

At the onset of the COVID-19 pandemic, the Commission asked platforms to report monthly on the effectiveness and impact of their policies in relation to the COVID-19 pandemic. These reports included detail on blocking accounts that engage in vaccine-related disinformation, advertisers disseminating COVID-19 and vaccine-related misinformation, and efforts to increase engagement with fact-checking organisations.

In January 2021, the European Commission provided some advice to assist signatories with their reporting. Signatories were advised to provide more data on the evolution and spread of disinformation during this time, the granular impact of their actions at the level of EU countries, and to have a greater focus on vaccine-specific disinformation. Monthly reporting requirements are expected to continue until at least the end of 2021.

Strengthening the EU Code

On 26 May 2021, the European Commission issued formal guidance setting out how platforms should step up their measures to address the shortcomings identified in the assessment of the EU Code.

Some of the key recommendations from the guidance included:

- > an expansion of the code to include misinformation in some areas, as well as private messaging services
- > wider participation from both established and emerging platforms, as well as fact-checkers, content assessment organisations, and technology developers
- > a requirement for signatories to publicly justify their reason for opting out of certain code provisions
- > enhanced code provisions on the demonetisation of disinformation on advertising channels.

It also suggests an enhanced monitoring and reporting framework, which includes the development of:

- > platform-specific and industry-wide KPIs
- > a harmonised reporting template to enable cross-platform comparisons
- > publicly accessible transparency centres created by platforms
- > a new data access framework for the research community
- > a dedicated taskforce consisting of representatives from signatories, EDMO, ERGA and other relevant experts, to evolve and adapt the code to technological, market, and legislative developments.

The Commission expects a first draft of a revised EU Code in the European autumn 2021.

Digital Services Act

The European Commission has proposed a Digital Services Act (DSA) to enable users, consumers, and businesses to continue to operate and grow with digital developments and challenges in Europe.²⁵² Its goal is to create a safer digital environment in which the fundamental rights of all users of digital services are protected. It is not anticipated that the DSA will be enacted before the end of 2022.

²⁵² European Commission, [The Digital Services Act package](#), December 2020.

The DSA provides a series of new, harmonised obligations for digital services in the EU. Some of these include:

- > rules for the removal of illegal goods, services, or content online
- > safeguards for users whose content has been deleted by platforms
- > obligations for platforms to take risk-based action to prevent abuse of their systems
- > transparency measures for online advertising and platform algorithms
- > facilitating access for researchers to key platform data, and public access to independent audit reports of very large online platforms' risk mitigation measures.

Services' obligations under the DSA are graduated on the basis of their size and reach in the EU. It is proposed that arrangements will be overseen by EU Member States, and for some larger platforms, the European Commission directly.

The DSA proposes establishing a co-regulatory backstop, which would include a strengthened EU Code and a more robust framework for monitoring the code.

The proposal also covers new advertising rules that will empower users in understanding and making informed decisions about the ads they see. Users will be clearly informed whether and why they are being targeted by each ad and who paid for the ad. They should also see clearly when content is sponsored or organically posted on a platform.

As a complementing piece to the measures proposed in the DSA, the European Democracy Action Plan was presented by the European Commission in December 2020 and contains a strategic plan specific for countering disinformation and its impact on democracy in Europe. The action plan proposes improving the existing EU's toolbox for countering foreign interference, such as new powers that allow costs to be imposed on bad actors. The European Commission will increase funding to support new innovative projects to fight disinformation and promote media literacy, including those by civil society organisations and academic institutions.

United Kingdom

Online Harms White Paper and draft online safety bill

The UK government's *Online Harms White Paper* puts forward a regulatory framework that assesses potentially harmful content in a proportionate manner and is based on evidence of risk of harm. The proposed approach aims to improve transparency for users about which content is and is not acceptable on different platforms and will enhance users' ability to challenge the removal of content where this occurs.

In December 2020, the UK government released its preliminary response of the consultation on the *Online Harms White Paper*.²⁵³ It appointed Ofcom as the independent regulator that will be responsible for protecting users from online harms.

The white paper has informed the development of the UK's May 2021 draft Online Safety Bill.²⁵⁴ The Bill applies to a range of user-to-user and search services that are categorised based on their online presence, functionality, and high-risk features.

²⁵³ UK Department for Digital, Culture, Media & Sport, [Online Harms White Paper – Initial consultation response](#), December 2020.

²⁵⁴ UK Department for Digital, Culture, Media & Sport, [Draft Online Safety Bill](#), May 2021.

Category 1 services, likely to include Facebook, TikTok, Instagram and Twitter, will need to mitigate the risk of harmful and illegal content on their services, including disinformation. They will also need to conduct and publish assessments of the steps they have taken to address any adverse effects caused by their platforms. Ofcom will publish guidance to assist providers to comply with their obligations to carry out risk assessments.

Category 1 services will also have a statutory duty of care to safeguard UK users' access to journalistic content shared on their platforms. Under this duty of care, services will need to consider the importance of journalism when undertaking content moderation, have a fast-track appeals process for journalists' removed content, and will be held accountable for the arbitrary removal of journalist content. Articles by recognised news publishers shared on Category 1 services will be exempt.

Ofcom will establish and maintain an advisory committee specific to addressing misinformation and disinformation and will advise on how:

- > regulated services should deal with such content on their services/encountered in or via search results
- > it will use its transparency reporting powers in relation to misinformation and disinformation
- > it will use its media literacy functions to counter such content.

The committee will publish a report within 18 months and maintain subsequent periodic reporting.

COVID-19 response

In March 2020, the Cabinet Office and Number 10 began implementing a specialist rapid response unit to combat false and misleading narratives about COVID-19, to ensure the public has access to credible and authoritative information to protect themselves.²⁵⁵

The government has also agreed on measures with Google, Facebook, and Twitter to limit the spread of disinformation and misinformation regarding COVID-19 vaccines.²⁵⁶

Together, these platforms have agreed to:

- > ensure a timely response to disinformation and misinformation content flagged to them by the government
- > continue to work with public health bodies to ensure that authoritative messages about vaccine safety reach as many people as possible
- > participate in policy forums to improve responses to disinformation and misinformation and prepare for future threats.

France

France adopted laws in late 2018 against the 'manipulation of information' (*Les enjeux de la loi contre la manipulation de l'information*), which apply to online platforms with more than 5 million monthly French users, including social media sites and search engines.²⁵⁷

²⁵⁵ UK Government, [Government cracks down on spread of false coronavirus information online](#), March 2020.

²⁵⁶ UK Department for Digital, Culture, Media & Sport, [Social media giants agree package of measure with UK Government to tackle vaccine disinformation](#), November 2020.

²⁵⁷ *Les enjeux de la loi contre la manipulation de l'information 2018*.

The laws encourage platforms for take measures such as:

- > improve transparency of algorithms
- > promote content from credible sources
- > fight against accounts that disseminate fake information
- > provide users with further insight into the source of content, especially advertisements
- > provide users with information on how the content is being disseminated
- > promote media literacy.

The national media regulator, Conseil supérieur de l'audiovisuel (CSA), has formal information-gathering powers to oversee digital platforms, including the ability to conduct performance reports and issue recommendations to platforms. It can also request certain data points from platforms and broader information requests such as providing more insight into the transparency of algorithms.

Platforms are required to submit annual performance reports to CSA outlining the measures they have taken to fight against disinformation in France. The first reports were supplied in May 2020 and CSA released its first evaluation report in July 2020.

As an EU Member State, France's national regulatory response is operated in parallel with the EU Code framework.

Germany

Germany also has national level arrangements to complement the EU Code. While there is no general law that prohibits the creation and dissemination of disinformation, there are several legal provisions that may be applicable to safeguard individuals or the public from disinformation on digital platforms.

In 2017, Germany passed the *Network Enforcement Act*, which aims to combat hate speech and misinformation on social media platforms that have more than 2 million registered users in Germany.²⁵⁸ Under the Act, social media platforms have 24 hours after receiving a user complaint to remove any content that is 'clearly illegal', as defined by provisions of the Criminal Code that cover the dissemination of propaganda, encouragement of serious violent offenses, the incitement of crime and hatred, among others.

Social media platforms must also offer their users a complaints mechanism that is accessible and transparent. Any decision made on a user's complaint must be communicated to the user and any affected user in a clear and timely manner.

Germany has also recently updated its interstate media treaty for broadcasting regulation. The treaty contains updated and expanded journalistic standards that will provide a legal basis for regulation of online services within its broadcasting framework. Article 19 of the Act seeks to protect free expression within the bounds of journalistic standards.

²⁵⁸ Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act, NetzDG) - Basic Information 2017.

Ireland

As an EU Member State, Ireland's response to disinformation has primarily fallen under the framework of the EU Code.

The Broadcasting Authority of Ireland (BAI) is the national regulatory authority for Ireland in relation to the EU Code, and has played a leading role in the ERGA subgroup on disinformation.

In April 2020, BAI published its report *CodeCheck: A Review of Platform Compliance with the EC Code of Practice on Disinformation* as its contribution to ERGA's assessment.²⁵⁹ The report outlined the responses by the platforms under Pillars A and C and examines in-depth specifically which of the actions occurred within Ireland under Pillar D and E. It found that significant progress was made by digital platforms. However, there remains weaknesses in the content of the code and procedures for reporting, monitoring, and enforcing the commitments, particularly at a national level.

The Irish Government is currently drafting the Online Safety and Media Regulation Bill. It is expected that the proposed legislation will address the regulation of harmful online content. BAI will be dissolved, and existing personnel transferred to a new Media Commission. An Online Safety Commissioner with specific responsibility for overseeing the regulatory framework for online safety will also be established within the Media Commission.²⁶⁰

United States

Under Section 230 of the US *Communications Decency Act*, providers and users of an 'interactive computer service' are provided a general immunity from civil liability arising from third-party content that is deemed obscene, excessively violent, harassing, or otherwise objectionable. Under this Act, digital platforms are treated as distributors – not the publisher or speaker – of any information provided by a user of the service.

There has been recent interest in whether actions taken by platforms against misleading content threatens freedom of expression. Under the former Trump administration, there were discussions about reforming Section 230, including an Executive Order on Preventing Online Censorship that sought to clarify whether platforms' actions constitute editorial conduct and therefore run against the principles underlying the immunity from liability.

The Biden administration has yet to announce a formal position on changes to section 230. However congressional hearings continue to examine the impact of the legislation. In March 2021, chief executives from Facebook, Alphabet (Google) and Twitter were asked to testify to the House Committee on Energy and Commerce relating to the role and responsibility of digital platforms in promoting extremism and misinformation following the 2020 US election.

Canada

Canada is undertaking ongoing efforts to modernise its legislative and regulatory framework for communications.

²⁵⁹ Broadcasting Authority of Ireland, [New report highlights inconsistencies across digital platforms in tackling disinformation](#), April 2020.

²⁶⁰ Mondaq, [Ireland: Irish Media and Broadcasting Law Update](#), May 2021.

The Canadian government's 2020 report *Canada's Communications Future: Time to Act* acknowledged the need to address online misinformation and recommended legislative reform to address the spread of harmful content on digital platforms.²⁶¹

During the COVID-19 pandemic, Canada's federal government has been underscoring the importance of trusting experts and consulting credible sources. It has invested significantly in projects that facilitate public awareness tools and online workshops to help citizens become more resilient and think critically about COVID-19 disinformation.²⁶²

The Canadian Department of Heritage has also proposed new measures to address online harms. Legislation would see the creation of a new regulatory body to implement rules for online speech and assess and enforce compliance with new regulations. The new regulator would be given auditing powers over platforms' content moderation and be able to implement a 24-hour takedown notice regime to help reduce the spread of potentially harmful content online.

The Canadian government is also leading a Diversity of Content Online Working Group with members from Australia, France, Germany, Finland, as well as from the private sector and civic society. In June 2021, a new set of guiding principles were announced to foster greater exposure to diverse news and information, bolster resilience to disinformation and misinformation, and promote greater transparency of the impacts of algorithmic treatments of online content.

The principles are intended to guide a range of stakeholder actions, including those of digital platforms. Signatories have agreed to develop specific commitments by December 2022 to implement the principles.

Singapore

The *Protection from Online Falsehoods Manipulation Act* (POFMA) allows government ministers to order news outlets, internet service providers, digital platforms and users to include warnings that their pages or posts contain false statements and include links to government fact-checking websites.²⁶³

The Infocomm Media Development Authority (IMDA), a statutory board under the Minister for Communications and Information, is responsible for administering POFMA and has invoked it more than 50 times, primarily against independent media or people who have criticised the government or its policies.²⁶⁴ The POFMA Office, situated within IMDA, issues directions and notices upon the instructions of ministers and monitors and enforces compliance with these directions.

Critics have said that POFMA is a threat to free speech and have highlighted the inherent conflicts of interests in laws that give the government extensive powers and discretion to decide what constitutes misinformation.²⁶⁵ Digital platforms have also raised concerns, saying that POFMA is a censorship tool, and activists and social

²⁶¹ Recommendation 94: 'We recommend that the federal government introduce legislation with respect to liability of digital providers for harmful content and conduct using digital technologies ... we also encourage the federal government to continue to participate actively in international fora and activities to develop international cooperative regulatory practices on harmful content'.

²⁶² Government of Canada, [Supporting Canadians to Think Critically About Online Health Information](#), April 2020.

²⁶³ POFMA Office, [Protection from Online Falsehoods and Manipulation Act \(POFMA\)](#), 2021.

²⁶⁴ Human Rights Watch, [Singapore: 'Fake News' Law Curtails Speech](#), 2021.

²⁶⁵ La Trobe University, [Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific](#), Carson A et al, 2021.

groups are fearful that it is being used for political gain. The Singapore government maintains that the law only tackles falsehoods and that legitimate criticism and free speech are not affected.

POFMA has also had a controversial impact on Singapore's online political advertising, with rules requiring internet intermediaries to keep records of all online political advertising content. This has prompted Google to ban all such content in Singapore.

Malaysia

Malaysia has recently adopted the Emergency (Essential Powers) (No. 2) Ordinance 2021. The ordinance makes it a criminal offense to create, publish and disseminate 'fake news' that is likely to cause fear or harm to the public; or to fail to take down such material upon the government's request. 'Fake news' is defined as 'news, information, data and reports which is or are wholly or partly false relating to COVID-19 or the proclamation of emergency'. It draws heavily from the repealed *Anti-Fake News Act 2018* and is effective until 1 August 2021.

The ordinance has been criticised for failing to establish standards for determining what is false, raising the risk that it could be used to silence criticism or other content that government may not like.²⁶⁶ It also allows for criminal punishment, regardless of whether the offending individual or company had a prior understanding of the content being false, misleading, or deceptive.

Taiwan

Similar to the EU code, major digital platforms such as Google, Facebook and Yahoo and other local industry groups in Taiwan have signed a voluntary, self-regulatory code to address concerns about false information on digital platforms.²⁶⁷ Much of the reporting framework from the EU has been translated into the Taiwan code, including platforms periodically reviewing the results of their activities and proactively continuing to establish dialogue with third parties and government agencies to support and maintain transparency.

India

In February 2021, the Information Technology (Guidelines for Intermediaries and Digital Media Ethics Code) Rules commenced. The rules are overseen by the Ministry of Electronics and Information Technology and apply, among others, to publishers of news and online curated content, and intermediaries that enable the transmission of such content.

Social media platforms will need to appoint a grievance officer who acknowledges complaints within 24 hours and resolves them within 15 days. They will also be required to set up a robust complaints-handling mechanism and publish a monthly compliance report detailing the complaints received and any subsequent action taken.

²⁶⁶ Human Rights Watch, [Malaysia: Revoke 'Fake News' Ordinance](#), March 2021.

²⁶⁷ Central News Agency (CNA), [Five major players such as Facebook, LINE to prevent false information take the lead in self-discipline](#) (translated), June 2019.

Certain platforms will be required to provide details about the origin of pieces of content and messages that has been identified as offensive by the Indian government. This extends to private messaging services that offer end-to-end encryption for their users. Some experts have expressed concern that this approach may suppress freedom of expression and increase censorship by digital platforms.

Appendix E: Other Australian Government initiatives

There are several government initiatives that intersect with the operation and administration of the code and its intended policy objectives. Many of these initiatives result from recommendations of the ACCC's Digital Platforms Inquiry (DPI).

This appendix provides public information about some of the initiatives currently underway.

Government responses to the Digital Platforms Inquiry

News Media and Digital Platforms Mandatory Bargaining Code

The news media bargaining code was developed following a recommendation of the ACCC's DPI. It received royal assent on 2 March 2021. The purpose of the news media bargaining code is to support the sustainability of the Australian news media sector by addressing the imbalance in bargaining power between digital platforms and Australian news businesses. It provides a framework for designated digital platforms and registered news business corporations to make commercial agreements regarding the availability of news on digital platform services. The obligations under the code are enlivened when a digital platform is designated by the Treasurer and the ACMA has registered a news business.

Under the news media bargaining code, designated platforms must develop proposals to recognise original news content when they make it available and distribute that content. This intersects with Objective 4 of the code, which places obligations on digital platforms to enable users to make more informed choices about the source of news and factual content accessed via their platforms.

On 3 March 2021, the ACMA opened applications for Australian news businesses to register under the news media bargaining code. At this stage, no digital platforms have been designated. Both Google and Facebook have struck commercial arrangements with a number of news outlets voluntarily following the passage of the legislation. The ACCC has also granted interim authorisation for Country Press Australia (CPA) members to collectively negotiate with Facebook and Google over payments for their news content that appears on these platforms.

Improving digital media literacy in the community

The Department of Infrastructure, Transport, Regional Development and Communications (DITRDC) is responsible for implementing Recommendation 12 of the DPI, to establish a network of experts to develop media literacy materials. DITRDC is focusing on adult media literacy, with particular attention to vulnerable and Culturally and Linguistically Diverse (CALD) communities. The work is in early stages of development, having been reprioritised due to the impacts of the COVID-19 pandemic. Currently, it is looking at opportunities to work with the already established Australian Media Literacy Alliance, which comprises a range of members, including the ABC, universities, the National Film and Sound Archive, Museum of Australian Democracy and Australian Library and Information Association.

Online Privacy Code and review of the Privacy Act

The Attorney-General's Department (AGD) is progressing reforms to strengthen privacy protections online by introducing a new binding online privacy code for social media and online platforms that trade in personal information and increasing penalties

and enforcement measures. The online privacy code will require organisations to be more transparent about the handling of personal information, stop using or disclosing an individual's personal information upon request, and follow stricter rules about handling personal information of children and other vulnerable groups. An exposure draft will be released shortly for public consultation.

As industry will be invited to develop the online privacy code following passage of the legislation, DIGI may have a role in code development. If an industry code developer cannot be identified, the new binding code would be developed by the OAIC.

Additionally, in response to recommendations in the DPI, AGD is conducting a review of the *Privacy Act 1988*. The AGD has been considering submissions received in response to its issues paper, which outlined the current law and sought feedback on potential issues relevant to reform. The AGD has also conducted targeted consultation with stakeholders, including other government departments and agencies, state and territory government departments, private sector entities, stakeholder representative organisations and peak bodies, and international governments and privacy regulators. A discussion paper analysing stakeholder feedback and seeking submissions on reform proposals is expected to be issued in 2021. Feedback received through submissions to that discussion paper, and further consultations will inform the review's final report for government. Some of the issues being considered by the review include whether the requirements for notice to collect personal information should be strengthened, whether individuals should be able to object to the collection of their personal information, and a potential right to erasure of personal information

External dispute resolution

Following Recommendations 22 and 23 of the DPI, DITRDC is leading a process to understand the existing dispute resolution mechanisms of digital platforms. The process involves mapping the internal dispute resolution practices of major digital platforms and gathering data from consumers and businesses to better understand their concerns. DITRDC is consulting with industry and government stakeholders to inform the development of policy advice.

There may be some similarities between this initiative and section 7.4 of the code, which obliges signatories to establish a facility for addressing non-compliance by signatories with the code.

ACCC digital platform services inquiry 2020–25

On 10 February 2020, the government, in response to the recommendation of the DPI, directed the ACCC to conduct an inquiry into markets for the supply of digital platform services, including search engine services, social media services, online private messaging services, digital content aggregation platform services, media referral services and electronic marketplace services. The ACCC must report to the Treasurer every 6 months until 31 March 2025.

The ACCC published its first interim report on 30 September 2020. The report examined online private messaging services in Australia and updated the ACCC's previous analysis of search and social media platforms. In this interim report, the ACCC observes that the terms and conditions of online private messaging services allow for the collection of a broad range of information about users, but do not provide clarity on how that data will be used. The ACCC also found that standard terms provided to small businesses seeking to advertise on large digital platforms are potentially unfair. The report continues to advocate for Recommendations 22 and 23 of the DPI regarding internal dispute resolution mechanisms and the establishment of an ombudsman scheme to resolve complaints and disputes with digital platforms.

The ACCC published its second interim report on 28 April 2021. This report provided in-depth consideration of competition and consumer issues associated with the distribution of mobile apps to users of smartphones and other mobile devices. Among other things, the ACCC found that Apple and Google could do more to prevent and remove apps that feature subscription traps and other scams. The report also found a need for better redress and dispute resolution for consumers harmed by such apps. The ACCC continues to support Recommendations 22 and 23 of the DPI. The ACCC considers that these mechanisms could cover complaints by both third-party app developers and mobile app users to help address identified deficiencies.

ACCC digital advertising services inquiry

On 10 February 2020, the government directed the ACCC to conduct an inquiry into markets for the supply of digital advertising technology services and digital advertising agency services. On 28 January 2021, the ACCC published its interim report. The interim report does not make any preliminary findings, but seeks feedback on a number of proposals, including that industry develop a voluntary standard to enable full, independent verification of services provided by demand-side platforms (for example, Google Ads and Display & Video 360). This proposal seeks to address opacity in the supply chain that prevents users from making informed decisions about the use of Google's advertising services.

The ACCC is due to provide a final report to the Treasurer by 31 August 2021.

Other government initiatives

Online Safety Act

The Online Safety Bill 2021 (Online Safety Bill), which passed the Senate on 22 June 2021 and is expected to receive royal assent in the coming weeks, provides for the development of a set of basic online safety expectations for social media services, relevant electronic services and designated internet services. The core expectations include that the provider of the service will:

- > take reasonable steps to ensure that end-users are able to use the service in a safe manner
- > provide reporting and complaints mechanisms for end-users
- > provide specified information to the eSafety Commissioner on request.

The Online Safety Bill also empowers the eSafety Commissioner to require the providers of these services to prepare either ad hoc or periodic transparency reports about their compliance with the basic online safety expectations during a specified timeframe.

In addition, the Online Safety Bill provides for the establishment of a complaints and take-down scheme for serious cyber-abuse material targeting Australian adults on social media services, relevant electronic services and designated internet services. This is material that an ordinary reasonable person would conclude is likely intended to have an effect of causing serious harm to an Australian adult, and which an ordinary reasonable Australian adult would regard as being menacing, harassing or offensive in all the circumstances.

The updated online content scheme within the Online Safety Bill provides for the removal of seriously harmful material in certain circumstances. It reflects and clarifies the current regime in Schedules 5 and 7 of the BSA. Under the scheme, the eSafety Commissioner's take-down powers for class 1 content (such as child sexual exploitation material and pro-terrorist material) are expanded to reach content that is hosted overseas. The eSafety Commissioner is also empowered to require services

provided from Australia restrict certain class 2 content (which maps to R18+ material under the National Classification code) for users under the age of 18. In addition, 8 sections of the online industry (including digital platforms represented by DIGI) are to work collaboratively to develop new industry codes to address class 1 and class 2 content. The Bill provides that the eSafety Commissioner should make reasonable efforts to ensure that an industry code is registered within 6 months of the Act commencing. The eSafety Commissioner would also have the power to create industry standards.

Defamation law reform

In November 2004, Attorneys-General endorsed the Model Defamation Provisions (MDPs). The states and territories then enacted legislation to implement the MDPs within their jurisdictions, collectively referred to as the National Uniform Defamation Law. A process to review Australia's National Uniform Defamation Law is currently underway. This review is being led by Attorneys-General and is being progressed in 2 stages.

Stage 1 is now largely complete, with Attorneys-General endorsing the Model Defamation Amendment Provisions (MDAPs) in June 2020. State and territory Attorneys-General have agreed that the MDAPs will commence within their jurisdictions on, or as soon as possible after, 1 July 2021. The Stage 2 review is underway and is on the 2021 agenda for the Meeting of Attorneys-General.

Part A of Stage 2 focuses on the responsibility and liabilities of digital platforms for defamatory content published online. Part A seeks to ensure that the MDPs are fit-for-purpose in the digital age, noting the 1 June 2020 decision by the NSW Court of Appeal to uphold a 2019 decision, which found news outlets were liable as 'publishers' for defamatory comments posted by third parties on their public Facebook pages. Part B of Stage 2 considers if defamation law is discouraging reports of misconduct to employers, police and other investigative or disciplinary bodies. Although the code does not address defamatory comments explicitly, Stage 2 defamation law reforms are consistent with a broader policy focus and regulatory push for digital platforms to be held more accountable for the content on their services.

Voluntary transparency reporting protocols

The Department of Home Affairs is leading Australia's engagement with industry, governments, academia and civil society to develop the Voluntary Transparency Reporting Framework under the auspices of the Organisation for Economic Cooperation and Development (OECD). The framework, co-funded by Australia, seeks to establish a common standard for online platforms to implement regular and transparent public reporting on the steps they are taking to prevent, detect and remove terrorist and violent extremist content on their platforms.

The third (and final) phase of the project commenced in May 2021 and is expected to be finalised later this year. Once established, the framework may be used as a standard of mandatory transparency reporting to the government for companies' responses to 'Abhorrent Violent Material' under the proposed Online Safety Bill. Australia is supportive of the OECD's plan to renew negotiations in 2022 to develop a more substantive Framework 2.0, which would be aimed at larger online platforms and services and will better reflect their capabilities and resources.

Department of Foreign Affairs and Trade ongoing counter-disinformation work

The Department of Foreign Affairs and Trade (DFAT) has established a counter-disinformation branch that works in collaboration with other agencies to monitor, analyse, assess and respond to disinformation that is contrary to Australia's national interests. It also engages with its overseas partners, including through the Australia-

United States Ministerial Consultations (AUSMIN) Working Group on Disinformation. A key element of this work is building resilience and limiting the spread of disinformation through capacity building in the Indo-Pacific region, and developing international norms to counter disinformation.

Home Affairs ongoing counter-disinformation work

In response to the threat of COVID-19 misinformation, disinformation and scams targeting Australians at a time of global crisis, Home Affairs established an All Source Fusion Cell (ASFC) to identify, analyse, assess and make recommendations for action to counter COVID-19 manipulated information activity and monitor key themes and trends in manipulated narratives.

The ASFC draws on information provided by departments and agencies across government and from open-source platforms to produce fused summary reports. Between March and June 2020, the ASFC produced over 60 reports, and made over 180 referrals to digital industry and law enforcement for further prevention, disruption and strategic communications action to minimise and counter the spread of misinformation and disinformation, and to protect Australians from scams.

Home Affairs has also been consulting with relevant departments and agencies to explore options for government to address malicious uses of the deepfake software, such as through the generation of disinformation, misinformation and malinformation.

Online Media Elections Protocol

The Electoral Council of Australia and New Zealand (ECANZ) is currently developing an Online Media Elections Protocol in consultation with online platforms. The protocol seeks to put in place arrangements between online platforms and electoral bodies at the state, territory or commonwealth level to deal with content that potentially breaches electoral legislation or the terms and conditions of online platforms.

The protocol makes references to relevant portions of the code. The code requires signatories to develop and implement policies that provide users with greater transparency about the source of political advertising carried on digital platforms.

Department of Health misinformation initiatives

The Department of Health (DoH) has a strategy to counter misinformation and works with Department of Home Affairs and state and territory departments, as well as a number of specific advisory groups (for example, the Aboriginal and Torres Strait Islander Advisory Group on COVID-19 and the Culturally and Linguistically Diverse Communities COVID-19 Health Advisory Group). As part of its strategy, DoH proactively engages with platforms via liaisons to promote reliable evidence-based information sources such as the DoH and World Health Organization.

DoH also monitors its own social media pages and accounts for misinformation and will report misinformation to digital platforms directly via the platform reporting functions or through platform liaisons. DoH also has social listening tools for uncovering broad trends, monitoring sentiment and emerging themes, and identifying popular hashtags. These social listening tools monitor external accounts to collect information, which is used to develop new content to address common themes.

Digital Economy Strategy 2030

On 6 May 2021, the Prime Minister, the Treasurer and the Minister for Superannuation, Financial Services and the Digital Economy released Australia's Digital Economy Strategy. The strategy sets out how Australia will secure its future as a modern and leading digital economy by 2030. It builds on the government's existing

data and digital initiatives, sets out further action under the 2021–22 Budget and defines future pathways to 2030.

The Digital Economy Strategy builds on the government's previous investments, including the Digital Business Plan, Cyber Security Strategy 2020 and Australia's Tech Future. Development of the strategy was led by the Digital Technology Taskforce in the Department of the Prime Minister and Cabinet, in close consultation across government and with industry, academia, and non-government organisations.

Australia's Cyber Security Strategy 2020

The Australia's Cyber Security Strategy 2020, developed by the Home Affairs in conjunction with PM&C, commits \$1.67 billion over 10 years to Australian cyber security initiatives. These include enhancing critical infrastructure and governmental networks, new laws to enhance privacy, consumer and data protection, blocking threats automatically and acting against cyber-attacks, as well as a voluntary code of practice for the Internet of Things (released on 30 September 2020).

Appendix F: Development of key performance indicators

The code is an outcomes-based, self-regulatory instrument. As articulated in the ACMA's position paper, an outcomes-based regulatory model has 3 distinguishing features:

1. Regulation is drafted as high-level outcomes or objectives that must be met.
2. Regulated entities develop their own systems to achieve the outcomes specified in the regulation.
3. Regulated entities are required to demonstrate delivery of these outcomes to the regulator, with enforcement and compliance measures in place should a failure to achieve an outcome occur.

The success of this approach hinges on signatories being able to demonstrate – both to DIGI as the code administrator and the wider community – how they are meeting each of their nominated outcomes under the code. Measures taken by signatories under the code should be capable of being monitored and measured over time, supported by a robust and transparent reporting framework.

As a requirement under the code, signatories have until 22 August 2021 to agree on a format for future annual reports and a guideline that will inform the data and other information to be included in subsequent reports. To promote consistency in reporting, we found that a uniform report format should be developed, requiring all code signatories to identify their services covered by the code, measures to address the outcomes and the specific metrics or KPIs they will use to measure success under each outcome.

This appendix outlines the ACMA's views and recommendations on measurement and reporting. Signatories may wish to consider this as guidance ahead of their next annual reporting process.

Measurement framework

KPIs are an important method to measure how effective an entity is performing against its objectives. KPIs need to be intimately connected to code commitments and be measurable through high-quality sets of data.

Data could be sourced through a variety of methods. The most valuable data is likely to come from signatories directly. Signatories may choose to use additional sources, such as consumer or academic research.

The ACMA has suggested that KPIs be separated into 2 tiers, based on whether they relate to the signatory's service(s), the industry, or the misinformation environment more broadly.

Tier 1 – Signatory-specific KPIs

A signatory publishes its annual report outlining how it will meet its commitments and what KPIs it will use to demonstrate performance under the code. These KPIs are monitored and reported on in each subsequent annual report.

The aim of signatory-specific KPIs is for signatories to demonstrate how their specific business will comply with the code. As each signatory can determine how they meet

their code obligations under an outcomes-based framework, signatory-specific reporting is necessary to demonstrate the performance of measures.

A signatory's annual reporting should cover both their progress towards implementing specific measures and the extent to which their measures have been successful in meeting the overarching objectives of the code. It is important that signatories are able to distinguish between the measures (for example a new tool or policy) and how they are to be assessed.

Examples:

- > A signatory has introduced a new tool that allows its users to view more details about who is funding advertising on its service. As part of its annual reporting, the signatory could report on the number of times this tool has been used by Australian users. This would include data directly from the signatory and include month-by-month increases or decreases in use.
- > A signatory has partnered with a local fact-checking organisation to independently review and tag content on their service. The signatory could report on the operation of the fact-checking organisation and provide quantitative data points such as how many pieces of content were tagged, how many accounts/pages and pieces of content tagged were Australia-specific, and the average time taken to review and tag content as false or misleading. Other data could include engagement with posts before and after being fact-checked.
- > A signatory has introduced a new tool that enables greater detection of bot activity. The signatory could report on any month-to-month or year-to-year changes in bot detection since the tool was introduced, how many bots discovered were being operated in Australia, or the average time for the signatory to identify and remove any bot accounts.
- > A signatory has implemented a new misinformation-related policy. It could outline why this policy is being implemented and demonstrate how it was communicated to users (such as through emails and/or notifications of changes to terms of service/users guidelines), and how the signatory will measure the success of this policy.

A signatory may have built a COVID-19 information and resource centre for its users to access during the pandemic. It could provide data on the number of Australian users who accessed this, a list of the official health links provided and the number of times Australian users clicked these links, and a list of the most common search terms used.

Tier 2 – Industry-wide and industry-group KPIs

DIGI, in collaboration with code signatories, should develop a series of cross-industry KPIs that align to the broader objectives of the code. This data allows for a better understanding of how each platform is performing under the code, as well as how signatories are performing as an industry in Australia.

Industry-wide and industry-group KPIs include broader metrics than signatory-specific KPIs and enable cross-platform comparisons.

These KPIs would need joint agreement by code signatories, be generalisable across all signatories and/or different groups of signatory-types (such as social media sites, search engines, or news aggregators), and directly align to the overarching objectives of the code. Given the breadth of signatories' service types, not all KPIs would apply to all signatories.

One of the main criticisms of the EU Code is the lack of KPIs to allow cross-platform comparison. As outlined in Appendix D, recent guidance from the European Commission provides that the EU Code should be strengthened with clear KPIs. This allows for a measure of both the impact of actions taken by signatories, as well as the overall impact of the EU Code on Disinformation in Europe. As no industry-wide KPIs were supplied in signatories' initial annual reports, we therefore strongly encourage the inclusion of these in the next code reporting process.

Signatories may choose to include these as part of their individual annual reports. Alternatively, they could report these metrics to an appropriate repository, such as DIGI or the code sub-committee, which could use this data in its oversight capacity. DIGI could report these metrics to the ACMA or publicly at an aggregated, industry-wide level so as not to call out any specific signatory.

Examples:

- > All signatories could agree to report annually on the number of Australian users who have made a report about harmful misinformation on their service(s) covered by the code, and the percentage of total reports that resulted in an action being taken by the platform. Signatories may provide high-level figures about whether the problem was addressed.
- > Signatories that offer social media services could demonstrate how they have addressed the same piece of content circulating on multiple platforms, such as a deepfake video or verifiably false information from a public figure. This could show similarities and differences in approaches and how efficiently certain content is addressed.
- > Signatories that offer search engine services could identify the number of sites and pieces of content that have been identified as disinformation and subsequently 'buried' in search algorithms.