Misinformation and news quality on digital platforms in Australia
A position paper to guide code development

JUNE 2020

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

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 or 1800 226 667
F +61 2 9334 7799

Contents

Contents (Continued)

Executive summary

There is significant concern about misinformation on digital platforms

Digital platforms are a key source of news and information for many Australians. However, Australians remain concerned about the accuracy and trustworthiness of news and information they consume online. Research from 2020 found that 48 per cent of Australians rely on online news or social media as their main source of news, but 64 per cent of Australians remain concerned about what is real or fake on the internet.[1]

Australians rely upon a range of indicators to assess the quality of their news and information, including the source or outlet of a news piece. On digital platforms, the widespread use of algorithms, the proliferation of sources and the dissociation of content from its source can make it challenging to assess quality and make informed decisions about which news and information to read and trust.

Difficulty in discerning the quality of news and information can lead to the increased spread of harmful misinformation. This includes disinformation—false and misleading information distributed by malicious actors with the intent to cause harm to individual users and the broader community.

International regulatory approaches to date have largely focused on countering deliberate disinformation campaigns.

Disinformation campaigns can engage ordinary users to inadvertently propagate misleading information. However, misleading information shared without intent to cause harm can still lead to significant harm. From the consumer perspective, all forms of false, misleading or deceptive information can have potentially harmful effects on users and the broader community.

This paper uses 'misinformation' as an umbrella term to cover all kinds of potentially harmful false, misleading or deceptive information, with deliberate disinformation campaigns considered a subset of misinformation.[2]

The government has been considering responses appropriate for Australian users

These concepts were canvassed as part of the Australian Competition and Consumer Commission (ACCC) Digital Platforms Inquiry (DPI). The ACCC recommended a mandatory code to address complaints about disinformation (Recommendation 15) and an oversight role for a regulator to monitor issues of misinformation and the quality of news and information (Recommendation 14).

In response to that inquiry, the government has asked major digital platforms to develop a voluntary code to cover both recommendations.

The government's response recognises that addressing the complex problem of misinformation requires a comprehensive and principled approach. Any such approach should balance interventions with the rights to freedom of speech and expression.

---

[1] S Park et al., *Digital News Report: Australia 2020*, News and Media Research Centre, University of Canberra, 2020, pp. 50, 77.

[2] The ACMA includes malinformation as an element of deliberate disinformation campaigns. Definitional issues are covered in detail in Chapter 2.

Australians are increasingly reliant on digital platforms to access, consume and share news and information. The ACMA considers that platforms bear considerable responsibility to provide users with a safe and user-friendly environment to engage with news and information and help users more easily discern the quality of this content.

The 2019–20 Australian bushfire season and the COVID-19 pandemic have reinforced the potential harms of false and misleading information

The first half of 2020 has been marked for many Australians by two extraordinary events: the unprecedented summer bushfire season and the COVID-19 pandemic. Both events have provided fertile circumstances for the spread of false and misleading information, distributed with and without malicious intent.

The bushfires saw instances of false and misleading information about the cause of the fires, the use of old images purporting to be of current events and conspiracy theories such as the fires having been purposely lit to make way for a Sydney to Melbourne train line.

False and misleading information about the pandemic—such as how to prevent exposure, possible treatments, and the origins of the virus—have been shown to have real-world consequences, including personal illness and damage to property. Recent Australian research found that nearly two-thirds (66 per cent) of people say they have encountered misinformation about COVID-19 on social media.[3] The World Health Organisation has labelled the crisis an 'infodemic' and platforms have implemented new measures to limit the spread of misinformation.

Both these events have highlighted the impact and potential harm of misinformation on both Australian users of digital platforms and the broader Australian community.

Voluntary codes should build on existing measures as part of a risk-based approach to harmful misinformation

In recent years, most major platforms have implemented a range of measures and processes to address potentially harmful misinformation and news quality issues. This work has intensified during the COVID-19 pandemic, with platforms taking further steps to address potential harms, including:

> Greater signalling of credible, relevant and authentic information through new features and tools.
> Increased detection and monitoring of fake accounts, bots and trolls who engage in malicious and inauthentic activity with vulnerable users.
> Updating terms of service and community guidelines to allow for action to be taken against false and misleading news and information in relation to health and safety issues where the scale and immediacy of potential harm is paramount.

In developing a voluntary code, the ACMA considers that platforms should codify their activities and commit to permanent actions that are systematic, transparent, certain and accountable for their users in addressing such potentially harmful misinformation.

A voluntary code needs to be fit for purpose for Australian users and the Australian community. Given the recent evidence of significant harm caused by false and misleading information shared online, and the practical difficulty of determining which information has been circulated with intent to harm, the ACMA considers platforms

---

3   S Park, et al., *COVID-19: Australian news and misinformation*, News and Media Research Centre, University of Canberra, 2020, p. 27.

should implement measures to address all kinds of harmful misinformation circulating on their services. These measures should be graduated and proportionate to the risk of harm.

Adopting a graduated and flexible approach means platforms would also be free to draw the lines between different interventions in accordance with their own policies and to achieve an appropriate balance with rights to freedom of speech and expression.

The ACMA has outlined its expectations to guide code development

This paper includes a series of positions that outline the ACMA's expectations on the development of the code. These positions cover threshold issues about the scope, design, and administration of the code, and are intended to assist platforms in the development of their code(s). These positions have been informed by existing international regulatory approaches, preliminary discussions with platforms and an examination of best-practice guidelines.

The ACMA considers that the code should cover misinformation across all types of news and information (including advertising and sponsored content) that:

> is of a public or semi-public nature

> is shared or distributed via a digital platform

> has the potential to cause harm to an individual, social group or the broader community.

To enable a consistent experience for Australians who use multiple platforms, the ACMA considers a single industry code would be the preferable approach. Any code should be consumer-centric, including providing a mechanism for users to easily access dispute resolution mechanisms.

As a voluntary code, it will be a matter for individual platforms to decide on whether they participate in the development of the code or choose to be bound by the code. The ACMA would, however, strongly encourage all digital platforms with a presence in Australia, regardless of their size, to sign up to an industry-wide code to demonstrate their commitment to addressing misinformation.

At a minimum, the code should apply to the full range of digital platforms that were outlined in the DPI terms of reference. This includes online search engines, social media platforms and other digital content aggregation services with at least one million monthly active users in Australia. The ACMA considers that this will likely include widely used platforms such as Facebook, YouTube, Twitter, Google Search and Google News, Instagram, TikTok, LinkedIn, Apple News and Snapchat.

The ACMA anticipates that code signatories will change over time to adjust to new entrants and other market changes.

The ACMA have developed a code model, using an outcomes-based approach, to assist platforms in composing their codes

In developing a code, the ACMA considers that platforms should adopt an outcomes-based approach. This would provide signatories with a common set of aims while granting the flexibility to implement measures that are most suited to their business models and technologies.

The ACMA has developed the code model below which articulates potential objectives and outcomes for the code.

Figure 1:   Code model: potential objectives and outcomes

PURPOSE	To provide Australian users of digital platforms with appropriate protections and remedies relating to misinformation		
OBJECTIVES	Reduce the impact of potentially harmful misinformation	Empower users to identify the quality of news and information	Strengthen transparency of and accountability for measures to combat misinformation
OUTCOMES	Users are less likely to be exposed to potentially harmful misinformation on the platform / Users are easily able to report potentially harmful misinformation on the platform	Users are better equipped to evaluate the quality of news and information on the platform / The source of political and issue-based advertising and sponsored content is transparent to users on the platform	Users are informed about what is being done to address misinformation on the platform / Users have access to a robust and effective complaints handling process under the code
MEASURES	To be determined by each platform based on the likelihood and severity of harm		

The ACMA considers that platforms should agree to implement a range of measures to address the proposed outcomes and objectives of the code. The measures should be graduated and proportionate, based on an assessment of the risk of harm and the unique characteristics of their platforms.

Platforms also need to consider how they will implement performance reporting arrangements to provide transparency around the implementation of their commitments and the operation of the code. Once a code is in place, the ACMA would encourage code signatories to publish individual action plans on how they will meet their commitments, as well as annual progress reports.

The ACMA will be developing a misinformation and news quality monitoring and reporting framework to inform its advice to government in June 2021

The government has requested that the ACMA provide a report by June 2021 on the code process and the state of disinformation in Australia. The ACMA anticipates this will include an assessment of:

> the code development process

> the content of the code(s) and resulting measures

> the state of disinformation in Australia.

To inform this assessment, the ACMA will be developing a reporting and monitoring framework. The framework will be informed by ongoing discussions and workshops with platforms over the coming months.

The government has stated that it will assess the success of the codes and consider the need for further regulatory action in 2021.

1. Context

Digital platforms, including social media sites, online news aggregators, and search engines, play a key role in disseminating news and information in the online environment. Users need to be equipped to critically engage with this content and should expect platforms to support and safeguard their user experience.

The ACCC's DPI found that 'consumers accessing news through digital platforms potentially risk exposure to unreliable news through the spread of disinformation, misinformation and malinformation'.[4] The ACCC recommended the development of a mandatory code for disinformation (Recommendation 15) and that broader issues of misinformation and news quality be monitored and assessed for the need for further regulation (Recommendation 14).

In December 2019, the Australian Government released its response to the DPI.[5] To address DPI recommendations 14 and 15, the government has asked the major digital platforms to develop a voluntary code (or codes) of conduct for disinformation and news quality. The code(s) should address concerns regarding disinformation and credibility signalling for news content. The code(s) should also outline what the platforms will do to tackle disinformation on their services and support the ability of Australians to discern the quality of news and information.

The government has asked the Australian Communications and Media Authority (ACMA) to oversee the development of the code(s). The ACMA is the independent statutory authority that administers Australia's telecommunications and broadcasting regulation, including the development of codes of practice.

The ACMA will report to the government on the adequacy of the platforms' measures and the broader impacts of disinformation, with the first report due no later than June 2021. The government will assess the success of any code(s) and consider the need for any further reform at that time.

Purpose

This position paper will assist digital platforms to develop their code(s) by:

> outlining the ACMA's understanding of the problem
> articulating the ACMA's expectations of the objectives, outcomes and scope of the code(s)
> setting out a high-level framework to measure the effectiveness of platforms' measures and the broader impacts of misinformation and disinformation
> proposing a timeline and next steps for the code development process.

The paper also looks at international regulatory approaches and the responses taken by digital platforms in Australia to date. The ACMA has also undertaken a review of published research and academic reports, and considered best practice approaches to self-regulation and outcomes-based regulation.

---

[4]   ACCC, *Digital Platforms Inquiry—Final Report*, p. 280.

[5]   Australian Government, *Government Response and Implementation Roadmap for the Digital Platforms Inquiry*, 2019, pp. 6–7.

The ACMA's positioning and expectations have been informed by discussions with digital platforms including Facebook, Twitter, Google, Microsoft, TikTok, Apple and Amazon, as well as DIGI, the digital industry association that advocates for the digital industry in Australia. The ACMA has also consulted with relevant government agencies, including the ACCC, as well as with international regulators such as Ofcom, the Broadcasting Authority of Ireland and the Canadian Radio-television and Telecommunications Commission. These discussions have significantly enhanced the ACMA's understanding of platforms' current practices concerning disinformation. The ACMA thanks representatives of the platforms, DIGI and fellow regulators for their contribution to its understanding and expectations.

2. Misinformation and the quality of online news and information

The internet has fundamentally changed the way information is produced and distributed. To anyone with access to an internet connection, a world of information is readily available: to search, navigate and discover.

We can also participate in the production and communication of information in ways that were previously not possible. Knowledge and information are now frequently published and disseminated by ordinary users, as much as by the media.

The internet has also provided new social spaces for people to connect with family, friends and those with common interests, without regard for geographic distance.

These changes have brought significant benefits, but they have also created a vastly expanded and unfiltered information environment. Unreliable and misleading content is at least as accessible as professionally edited news or information published by authoritative sources.

Digital platforms and the online information environment

Digital platforms help us navigate the expanded online information environment by collating, filtering and prioritising information.

However, the same features of digital platforms that provide these benefits can also make it difficult for people to judge the reliability of the news and information they engage with online. These features can serve to amplify unreliable and deceptive information, and may be abused by malicious actors seeking to mislead or deceive.

The features of digital platforms and their effects

> *Expansion of public discourse*—the internet has expanded the realm of public discourse well beyond the bounds of professional journalism and information access and production are now both universal and global. Information is no longer necessarily mediated by the news media but can be published directly by governments, corporations, organisations, and individuals. This proliferation of sources means that unreliable information is as accessible as professional news. In addition, users have also become publishers and distributors, sharing and framing messages for their online networks. Much of the editorial burden of assessing the quality of information has shifted to consumers.

> *Information transparency*—online content is atomised, meaning that information is delivered piecemeal from multiple sources. This flattens the content, obscuring the indicators that consumers can use to help judge the reliability of news and information. It also reduces the editorial control of publishers, and users may not be exposed to the full editorial narrative that, over time, may include fact-checking or the coverage of alternative points of view. In addition, false and misleading information surfaces to users through the same channels as genuine news and information.

> *Algorithmic newsfeeds*—the use of algorithms to deliver content to users means that users do not have direct control over the content that they see. While algorithms automate content selection, platforms control the algorithms themselves, and make regular adjustments to them to favour certain inputs over others. These inputs do not necessarily parallel those used by news media in

making editorial decisions. They may include promoting content more likely to engage users and responding to a user's search or engagement history, which may reinforce the consumption of particular sources or points of view. Online, algorithms have taken over some of the editorial control traditionally held by the news media. In turn, news media have responded by producing content more likely to engage users and be promoted by algorithms.

> *Amplification*—content can spread very rapidly through user networks via conscious sharing and algorithmically driven amplification of trending stories. Unreliable or deceptive information often spreads faster than genuine news as it can appeal to emotion, curiosity or strongly held beliefs.[6] The volume of online content means that credible and authoritative information is in danger of being drowned out by less reliable information designed to attract attention.

> *Social engagement*—users participate in social media platforms not only to inform themselves but as a form of social engagement. They share, comment and like to show approval, argue, gain attention and entertain. Misleading or deceptive information often takes hold in communities of like-minded users and orchestrated campaigns can take advantage of this.

> *User targeting*—harvesting of user data and the prevalence of niche online communities enables advertisements, sponsored content and non-commercial messages to be targeted at carefully defined user cohorts, often with little transparency to the user or to the public. Messages may be sent from, and targeted at, users anywhere in the world. This problem may be exacerbated where news and information are delivered on private messaging services, potentially reducing the transparency of information and source.

> *Replication*—sharing generates many copies of the same content which may not be captured by efforts to remove or downgrade it in user feeds.

Given our increasing reliance on online platforms for news and information, and the benefits that platforms derive from this reliance, these platforms bear considerable responsibility to improve the online information environment. This includes helping users to more easily discern the quality of news and information, and reducing exposure to, and the impact of, false, misleading or deceptive information.

The problem of the online information environment is twofold:

1. It can be difficult for users to discern the quality of news and information online.

2. People are frequently exposed to false, misleading or deceptive information online—either through deliberate targeting by bad actors or ordinary users sharing information within their online networks.

Since each of these elements contributes to and exacerbates the other, an effective solution needs to address both these elements in tandem.

The quality of online news and information

Over time, the professional news media has developed a set of practices that serve to maintain quality, help the industry perform its social role of providing reliable information and provide a forum for debate about issues of public importance. In many countries, these practices have been formalised in codes of practice or journalistic conduct. These codes typically include provisions designed to encourage the supply of news and journalism that is accurate, transparent, fair, balanced or impartial.

---

6   C Wardle & H Derakhshan, *Information Disorder: Toward an interdisciplinary framework for research and policymaking*, Council of Europe, 2017.

Even within the parameters of professional news, the concept of news quality can be subjective, with consumers differing significantly in their preferences for, and judgements on the quality of, news sources. Given these differences, it is challenging to develop an agreed set of criteria or a single definition as the basis for an objective measure of news quality.

Consumers often rely on a range of common indicators when making these judgements. These include content indicators such as originality, relevance and timeliness, through to grammar, spelling and editorial style, as well as images, headlines and sources used. They may also include contextual information such as the publisher and particular author and the date and time of publication.[7]

As discussed in the previous section, many of these indicators are not as clear to users in the online information environment. This environment also includes a vast array of sources beyond professional news publishers and journalists.

A focus of platforms should be on increasing transparency in the online information environment—by, for example, increasing visibility of these indicators so that users are more easily able to make informed judgements about the quality of news and information they encounter. Many platforms have taken steps in this direction. Further codification would promote consistency and provide accountability for platforms in the services they provide to users and in their crucial role in the delivery and distribution of news and information.

However, the proliferation of sources beyond the traditional news media means that any scoping of this issue needs to include a broader range of online information than professionally produced news and journalism.

> For the purposes of this paper, the ACMA defines quality news and information as news and information that is accurate, reliable and timely, providing people with the knowledge they need to make informed choices and to participate in public life.

Addressing false, misleading or deceptive information

Challenges in identifying quality news and information on digital platforms are one of the factors that have increased people's exposure to misleading or deceptive information online.

Online, the burden of judging the quality of news and information is shifted considerably onto users. Malicious actors take advantage of this shift, making use of deceptive techniques to promote false and misleading content in coordinated disinformation campaigns. However, misleading content can also be spread by ordinary users amongst their online networks.

A common approach to understanding this problem defines three forms of misleading or deceptive information—or 'information disorder':[8]

> **Disinformation** is false or inaccurate information that is deliberately created and spread to harm a person, social group, organisation or country.

---

7   D Wilding, et al., _The Impact of Digital Platforms on News and Journalistic Content_, Centre for Media Transition, University of Technology Sydney, 2018, pp. 83–87.

8   The terminology is adapted from C Wardle & H Derakhshan, _Information Disorder: Toward an interdisciplinary framework for research and policymaking_, Council of Europe, 2017.

> **Misinformation** is false or inaccurate information that is not created with the intention of causing harm.
> **Malinformation** is accurate information inappropriately spread by bad-faith actors with the intent to cause harm, particularly to the operation of democratic processes.

Regulatory responses to information disorder, including that recommended in the DPI, have typically focused on disinformation due to the rise of coordinated online campaigns to interfere in elections or to undermine the exchange of reliable information that underpins democratic processes.[9] Less attention has been given to misinformation, in part due to legitimate concerns about freedom of expression and who should be the arbiter of truth. Malinformation is largely a feature of coordinated disinformation campaigns rather than an independent phenomenon.

In practice, it is difficult to distinguish between these types of information disorder, both for users and for platforms.

From the user perspective, all are forms of misleading or deceptive information and all can have potentially harmful effects on individuals and the broader community. Importantly, the potential for harm is not always tied to an intention to cause harm.

Recent misinformation crises such as the 2019–20 Australian bushfires and the COVID-19 pandemic have shown that the widespread sharing of inaccurate or misleading information has the potential to cause significant harm.

For platforms operating in real time, it can be difficult to determine the intent behind the sharing of misleading information.[10] The line between bad actors and innocent users is not always clear. Malicious disinformation campaigns work by engaging ordinary users to unwittingly propagate misleading information. They often begin by targeting online interest groups to co-opt ordinary users and encourage particular narratives. Authentic users, for their part, have varied and frequently indeterminate motivations for sharing misinformation.[11]

What renders information misleading or deceptive is also complex and context dependent. Much information disseminated maliciously is not demonstrably false or fabricated. It is 'often genuine, used out of context and weaponized by people who know that falsehoods based on a kernel of truth are more likely to be believed and shared'.[12] Some malicious actors deliberately spread contradictory messages to sow confusion and destabilise public discourse.

'Fake news' is another term that has been used to describe misinformation over recent years. Its meaning is elusive, and arguably encourages a narrow focus on news at the expense of the full array of misleading or deceptive information that circulates online. As First Draft's Clare Wardle points out, most misinformation or disinformation is not news, and doesn't even masquerade as news. Instead it's 'good old-fashioned rumours, it's memes, it's manipulated videos and hyper-targeted 'dark ads' and old photos re-shared as new'.[13]

---

[9]   Regulatory responses will be examined in Chapter 3.

[10] This is also a problem for researchers and analysts examining misinformation after the fact. See JS Brennen et al., *Types, Sources, and Claims of COVID-19 Misinformation*, Reuters Institute for the Study of Journalism, University of Oxford, 2020, p. 2, n1.

[11] C Wardle & H Derakhshan, *Information Disorder: Towards an interdisciplinary framework for research and policymaking*, Council of Europe, 2017, p. 23; JS Brennen et al., *Types, Sources, and Claims of COVID-19 Misinformation*, Reuters Institute for the Study of Journalism, University of Oxford, 2020, p. 5.

[12] C Wardle, *First Draft's Essential Guide to Understanding Information Disorder*, First Draft, 2019, p. 6.

[13] ibid.

For the purpose of devising an effective code of practice, it is important to take into account the practical difficulties of distinguishing between disinformation, misinformation and malinformation. It is also important to acknowledge the range of content which can be misleading or deceptive.

> For the purposes of this paper, the ACMA uses **misinformation** as an umbrella term to cover all kinds of potentially harmful false, misleading or deceptive information, with deliberate disinformation (and malinformation) campaigns considered a subset of misinformation.[14]

Harms caused by misinformation

Misinformation causes a range of harms to individuals and to society. These may be classified as acute and chronic harms. Examples of such harms (not intended to be exhaustive) are outlined below.

Acute harms Acute harms are harms that have an immediate impact on people, property or society.	
	Health and safety: misinformation about medical matters may directly threaten the health of anyone who acts on the information, and in the case of infectious disease, the health of others and public health systems. Misinformation may also threaten personal safety, as when a person is falsely identified on social media as the perpetrator of a crime, or by leading someone to act imprudently during a bushfire. It may also lead to abuse, harassment, or violence against a particular demographic group.
	Public panic and social disruption: misinformation may cause unrest or panic amongst a group of people or society more broadly. This may threaten the operation of public services and lead to harms to public or private property.
	Electoral integrity: misinformation about electoral rights, obligations or processes may threaten the integrity of an election.
	Financial and economic: misinformation may cause financial harm to individuals through scams or by influencing personal financial decisions. Misinformation may also cause economic harm via the disruption of business operations or financial systems.

---

14   This is the approach taken by JS Brennen et al., _Types, Sources, and Claims of COVID-19 Misinformation_, Reuters Institute for the Study of Journalism, University of Oxford, 2020.

Chronic harms Chronic harms are harms that result from the cumulative effect of misinformation, and may only become apparent over time.	
	Trust in public institutions: misinformation can, over time, negatively impact the public's trust in governments, legal systems and democratic institutions. As trust is essential for effective governance, this can lead to poor public policy outcomes and reduced engagement with mainstream democratic processes.
	Trust in professional sources of information: misinformation can, over time, erode the public's trust in professional expertise or sources of information, such as those originating from healthcare professionals, the media, academia or the scientific community. This can reduce citizens' ability to recognise facts and make informed decisions, while increasing their susceptibility to more acute harms from misinformation.
	Community cohesion: misinformation lowers the quality of public discourse and can lead to increased polarisation and poorer community cohesion. Misinformation can be designed to target certain groups, ethnicities or people with certain beliefs within society, and can result in reduced tolerance of these different groups or those with different opinions.

The need to address misinformation with a high risk of acute harm does not imply that this is the only misinformation over which platforms should be taking action. Although responses to chronic harms may be less urgent, the harms are not necessarily less severe. Societies around the world are grappling with coordinated campaigns designed to sow confusion and distrust and to undermine democratic institutions over time. There is a substantial risk to Australian society and its security if misinformation contributing to chronic harms is not adequately addressed.

Assessing the risk of harm

To determine what constitutes an appropriate response to misinformation, an assessment should be made of the overall risk of harm to users and the general public, in both the short and long term.

An effective approach to tackling information disorder will necessarily encompass harmful misinformation as well as coordinated disinformation campaigns.

This does not mean the measures to address misinformation and disinformation must be the same. Adopting a graduated, risk-based approach will give platforms the flexibility to implement measures tailored to their own services and business models.

This approach may also alleviate concerns about freedom of speech and who should act as the arbiter of truth. Including misinformation in the scope of the code does not mean that platforms should be expected to remove content merely because it is false or misleading. Instead, responses to misinformation should be proportionate to the risk of harm, with stronger measures such as removal applied to material or behaviour that presents a higher risk of harm to the public.

The principle of free expression is an essential consideration in the assessment of misinformation and may result in a higher bar for some content such as expressions of political opinion. Many platforms make use of third-party fact-checkers and other independent organisations to assist in content decisions or help set appropriate policies on misinformation.

Key factors in assessing risk of harm

There is a range of relevant factors to consider in assessing the overall risk of harm from misinformation. While not exhaustive, these include:

> *Content and context*—The subject and content of misinformation may be important in determining the risk of harm. Misleading health information, for example, may present a higher risk. Similarly, misleading safety information, such as that circulated during the 2019–20 Australian bushfire season, can cause confusion and misdirection within communities and put lives at risk. The context in which the information appears may also be important: health and safety misinformation during a pandemic or natural disaster is likely to cause greater harm than it otherwise might.

> *Agent and purpose*—Identifying who is spreading misinformation and why they are doing so may also be key factors in assessing the risk of harm. Bad actors launching a coordinated disinformation campaign are likely to persist, have multiple points of attack, be informed about their targets and what is likely to achieve their goals.

> *Breadth and speed of dissemination*—Understanding the breadth and speed of dissemination may be a key factor in assessing the risk of harm. A piece of information that is assessed as potentially harmful because of its content may not be widely shared, effectively lowering the risk of harm. Conversely, a seemingly innocuous item may be widely and rapidly shared, requiring a reassessment of the risk it poses.

> *Distribution channel*—Certain distribution channels may increase the risk of harm by allowing malicious actors to target particular users or groups that, for example, they consider more likely to spread the misinformation amongst wider networks. These could include private channels or messaging services.

The spread of misinformation during COVID-19

The COVID-19 pandemic has highlighted the complex and rapidly evolving nature of online misinformation, and the range of resulting acute and chronic harms. Labelled by the World Health Organisation as an 'infodemic'[15], it has demonstrated that:

> Malicious campaigns by state actors and scammers are only part of the problem and misinformation spread by ordinary users presents a substantial risk of harm.

> Online misinformation can have real-world impacts—with COVID-19 misinformation reported to be related to deaths and harm to individuals' health, damage to property, and erosion of trust in governments, media and public health initiatives.

Platforms have shown agility and willingness to introduce new measures in response to the spread of COVID-19 misinformation online. These include:

> updating policies, community standards or terms of use to include provisions on misinformation

> working with governments, other authorities and external partners to promote official and reliable sources and counter misinformation

---

[15] World Health Organisation, *Coronavirus disease 2019 (COVID-19) situation report – 86*, 15 April 2020.

> taking steps to reduce the vast amount of misinformation circulating on their platforms, for example by removing material with the potential to cause significant harm, labelling fact-checked material and providing contextual information from authoritative sources.

This sustained and multi-pronged response shows that platforms, public authorities, independent fact-checkers, news media and users all have a role to play.

While the COVID-19 pandemic is an extreme and unprecedented event, it provides an opportunity to better understand the impacts of misinformation and the mechanisms that platforms can employ to manage misinformation. Lessons from this 'infodemic' can inform approaches going forward to enable more effective and robust responses in the future.

A full case study is included at Appendix A.

3. Responses to misinformation

A range of international regulatory approaches and responses from platforms have informed the ACMA's positions in this paper.

International regulatory responses

Several countries have introduced or are considering anti-misinformation initiatives relating to digital platforms.[16] These may provide a helpful indication of what could be effective in an Australian context and provide a basis for a common approach across countries, particularly where those initiatives have now been in operation for some time and there has been some assessment of their effectiveness.

The EU Code of Practice on Disinformation (EU Code)

The EU Code is a non-binding, voluntary code that sets out a list of high-level commitments and principles that signatories agree to follow to protect users from disinformation. It was signed by Facebook, Google, Twitter, Mozilla and several advertising industry groups in October 2018, by Microsoft in May 2019, and by TikTok in June 2020. Upon signing, all signatories presented the European Commission with a roadmap of their proposed code implementation.

The scope of the EU Code is restricted to disinformation[17] and therefore it would not cover the full breadth of potential harms caused by misinformation during the 2019–20 Australian bushfire season and the outbreak of the COVID-19 pandemic. The disadvantages of this approach have been discussed in Chapter 2.

The EU Code outlines a range of commitments that signatories can choose to meet, including:

> greater scrutiny of ad placements, including disrupting advertising revenues of certain accounts and websites that spread disinformation

> making political and issue-based advertising more transparent

> addressing fake accounts and online bots

> empowering consumers to reporting disinformation and access different news sources, while improving the visibility and findability of authoritative content

> empowering the research community to monitor online disinformation through privacy-compliant access to the platforms' data.

One of the key strengths of the EU Code is that it is a single code that is applied equally to all signatories. This has enabled coordination and collaboration from signatories and a clear, systematic approach to implementation. It also allows for the European Commission to conduct a stronger comparative analysis of platforms' activities and reporting. It is encouraging to see that all leading industry organisations, including Facebook, Google and Twitter, have chosen to join the code.

As an outcomes-based approach to regulation, there is also value in the code being flexible in its approach and encouraging platforms to commit to finding their own solutions to the problem. Particularly in the dynamic and fast-evolving digital sector,

---

[16] This paper will highlight responses made in Europe, Taiwan and the United Kingdom. The ACMA has also considered other international responses, such as Italy, Singapore and Russia.

[17] Disinformation is defined in the code as 'verifiably false or misleading information', which, cumulatively, 'is created, presented and disseminated for economic gain or to intentionally deceive the public' and 'may cause public harm', intended as 'threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security'.

there is a need for adaptable responses to emerging new forms of disinformation and methods from bad actors. This allows platforms to address the same core principles and commitments in a manner that best suits their business model and capacity.

Signatories were required to identify the commitments they will adhere to, the specific policies and actions they will pursue to implement these commitments, current company best practices, milestones of the overall implementation of the EU code, and an annual account of their work. This has enabled the European Commission to carry out targeted monitoring and regularly meet with platforms to discuss their progress in implementing the code. Signatories have also regularly met to review and analyse their progress, implementation and effectiveness.[18]

There has been some criticism of the platforms' first annual reports. Both the European Commission and the European Regulators Group for Audiovisual Media Services (ERGA)[19] have expressed concerns about the failure of platforms to provide the necessary data and detail on the actions they have undertaken (particularly at a national level). They have also highlighted the lack of adequate benchmarks to evaluate progress and specificity showing that new policies and tools are working effectively.[20]

Concerns have also extended to the EU Code itself. The Commission's Independent Sounding Committee Board[21] and expert commentators[22] have argued that the code contains no clear and meaningful commitments, no measurable objectives or key performance indicators to conduct proper monitoring, and no compliance or enforcement tool.

For example, the code contains a commitment on empowering the research community, which includes providing researchers with access to data sets or collaborating with academics and civil society organisations in other ways. It does not, however, state exactly how this cooperation should work, the procedures for granting access and for what types of data, or which measures researchers should put in place to enable appropriate data storage and security. Three European Commissioners have warned that the access to data provided so far 'still does not correspond to the needs of the independent researchers'.[23]

ERGA and industry researchers have also pointed to the lack of transparency of the platforms' internal operations and decision-making processes, arguing that unless platforms are more amenable to thorough public auditing, a robust evaluation of the effectiveness of their initiatives is not feasible.[24] In addition, they argued that the EU Code needs a greater articulation of certain definitions (such as 'political advertising' and 'issue-based advertising') and its measurements and reporting structures, including stronger mechanisms through which self-assessment reporting can be verified.

---

18 *EU Code of Practice on Disinformation*, p. 9.

19 European Regulators Group for Audiovisual Media Services, *ERGA Reporting on Disinformation: Assessment of the Implementation of the Code of Practice*, 2020.

20 European Commission, *Code of Practice on Disinformation one year on: online platforms submit self-assessment reports*, 2019.

21 Sounding Board of the Multistakeholder Forum on Disinformation Online, *The Sounding Board's Unanimous Final Opinion on the So-Called Code of Practice*, 2018.

22 Carnegie Endowment for International Peace, *EU Code of Practice on Disinformation: Briefing Note for the New European Commission*, 2020.

23 European Commission, *Code of Practice on Disinformation one year on: online platforms submit self-assessment reports*, 2019.

24 Carnegie Endowment for International Peace, *EU Code of Practice on Disinformation: Briefing Note for the New European Commission*, 2020.

The European Commission is currently carrying out its assessment of the effectiveness of the EU Code, considering the self-assessment by the signatories and input from other key stakeholders. The report is due in mid-2020.

While the ACMA understands that progress has been made since platforms' self-assessment reports, these observations highlight the critical importance of sufficient data and information supplied by platforms. This will be vital to monitor the effectiveness of an Australian code.

Taiwan Code

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.[25]

Taiwan has also adopted a series of outcome-based commitments so that signatories can undertake different approaches to achieving a common outcome. 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.

United Kingdom's Online Harms White Paper

The UK government's *Online Harms White Paper* sets out a potential model to tackle the problems of disinformation by examining content through a list of harms.[26] While the paper acknowledges differences between misinformation and disinformation, it ultimately notes that any 'inaccurate information' can be harmful and that code commitments should aim to reduce such harm.

This harms-based approach is strongly focused on the user and their experience in the online environment. It supports a free, open and secure internet while simultaneously protecting freedom of expression and making people safer online.

The paper proposes establishing a statutory duty of care on relevant companies that requires them to take proportionate and proactive steps to keep their users safe and to tackle potential harm caused by content or activity on their services.

It contains many commitments similar to those detailed in in the EU Code but with a greater focus on user experience, policed through the creation of an independent regulator that has responsibility for protecting users from such harm. This approach has significantly informed the ACMA's expectations.

The white paper also states that the proposal put forward in the Cairncross Review[27] for a 'news quality obligation' to be imposed on social media companies is 'very much in line with our aim to strengthen the online environment'[28]. This obligation would require these platforms to improve how their users understand the origin of a news article and the trustworthiness of its source. The UK government is currently reviewing feedback received on the white paper.

---

25   Central News Agency (CNA), *Five major players such as Facebook, LINE to prevent false information take the lead in self-discipline* (translated), 2019.

26   UK Department for Digital, Culture, Media & Sport, *Online Harms White Paper*, 2019.

27   The Cairncross Review is a 2019 report into the sustainability and quality of the journalism industry in the UK, including the impact of search engines and social media sites, and the role of advertising.

28   UK Department for Digital, Culture, Media & Sport, *Online Harms White Paper*, 2019.

Responses from platforms

Digital platforms have implemented a range of measures to address misinformation and news quality. These include measures in response to international regulatory arrangements (as above), proactive and self-initiated measures to address real-time events, and developments in research and new technologies. Platforms have also launched collaborative initiatives with other stakeholders, including fact-checkers, academics, civil society and national authorities.

Many digital platforms have distinguished between measures addressing user behaviour (in particular user misconduct such as fake accounts, bots and trolls) and measures which address problematic content. Some platforms have advised that they currently focus on user behaviour rather than problematic content as a strategy to avoid internal bias and being seen as the arbiter of truth or news quality, particularly in political communication.

In high-level summary, platforms have reported broad efforts to:

> disrupt advertising and monetisation incentives that contribute to the dissemination of misinformation, such as fraudulent charities and scams

> increase transparency of advertisements and sponsored content, including clearly labelling such content to help user distinguishability

> increase detection and monitoring of fake accounts, bots and trolls, who engage in malicious and inauthentic activity with vulnerable users, including the use of artificial intelligence detection technology

> update terms of service and community guidelines to require users not to misrepresent their identity on platform service

> supply tools for users to report suspicious, fake or spam accounts on some digital platforms

> invest in technological means to signal credible, relevant and authentic information as well as invest in features and tools that make it easier for people to find diverse perspectives about topics of public interest, and

> provide researchers and the fact-checking community with financial support and access to data. Some platforms have direct contracts with local fact-checking organisations to vet and flag content posted on their services.

More recently, additional specific references to misinformation and the removal of false and misleading news and information have been added into some platforms' community standards and user guidelines, both generally and in relation to specific matters such as health and safety issues where the scale and immediacy of potential harm is paramount.

Further steps needed

The ACCC's DPI final report highlights that there are still significant and ongoing threats posed by misinformation and it remains prevalent on digital platforms.[29] Evidence also suggests that Australian consumers are very concerned about the extent of unreliable news.

---

29   ACCC, *Digital Platforms Inquiry - Final Report*, 2019, p. 354.

While platforms have certainly made progress in addressing misinformation and news quality—most recently and significantly during the COVID-19 pandemic—there is a need to build on and codify these initiatives so platforms' measures reflect user expectations and employ a consistent and robust approach to significant actual, potential or emerging harms. Codification would also assist new or growing industry participants to understand and adopt best practices.

Digital platforms have already implemented a range of measures to address misinformation and news quality on their service, including proactive measures and responses to other regulatory arrangements. It will be important for platforms to build on and codify these initiatives to meet user expectations and have a consistent, robust set of best practices for these issues.

4. ACMA positions on code development

Based on the ACMA's consideration of existing international regulatory approaches, preliminary discussions with platforms and an examination of best practice guidelines, the ACMA has formed a series of positions on the preferred scope, design and administration of the Australian code(s) of practice for disinformation and news quality.

By outlining these positions below, this chapter is designed to assist platforms in better understanding the ACMA's overarching expectations of the voluntary industry code(s) to provide Australian users of digital platforms with appropriate protections and remedies related to misinformation.

Scope of the code

Key code characteristics

The ACMA encourages all interested platforms to consider collaborating on, and contributing towards, the development of a single, industry-wide code[30] that can apply across the digital platforms market. Noting that the majority of Australians are active users of more than one platform, this would provide the Australian public with greater clarity about the nature of the code and greater certainty in engaging with online content more generally.

While platforms may choose to develop their own self-regulatory schemes, a single industry-wide code would also benefit industry in promoting a consistent and collaborative approach to dealing with misinformation across platforms, while providing efficiencies through standardised administration, complaints handling, reporting and measurement processes.

The ACMA expects a key characteristic of this code will be its consumer-centric focus. It follows that actions under the code would be framed in terms of the protections and remedies provided to Australian users of digital platforms, with the code itself drafted in plain English and presented in a way that is readily accessible to the general public.

The code should also be tailored to respond to the problems of misinformation in Australia and to address the potential harms to Australian users. While the ACMA would encourage platforms to build on successful self-regulatory schemes that have been adopted in other jurisdictions, efforts to simply rebadge existing codes without consideration of the local environment would not be considered an adequate response.

> **Position 1:** The ACMA encourages platforms to consider a single, industry-wide code that provides appropriate protections and remedies for Australian users of digital platforms. It expects this code will be consumer centric, readily accessible to the public, and fit-for-purpose for Australia.

Code coverage

The ACMA expects that the scope of the code will cover both misinformation and disinformation as addressed by recommendations 14 and 15 in the DPI final report. Given the experience of both the COVID-19 pandemic and the summer bushfire season, the ACMA considers a focus on disinformation to be too narrow for platforms

---

[30] For the purposes of this chapter, codes will be referred to in the singular, reflecting the ACMA's preference for a single code.

to adequately address the wide range of potential harms, including from content that has been distributed by those who genuinely believe it to be true and have no intent to cause harm.

For the purpose of devising an effective code of practice, it is important to take into account the practical difficulties of identifying an intent to harm. It is also important to acknowledge the range of content which can be misleading or deceptive. An outcomes-based approach will give platforms the flexibility to determine what measures they employ to address different forms of misinformation on their services.

As specified in the government response to the DPI, the code should cover online information as well as news and journalism. The ACMA is of the view that the code should cover misinformation across all types of news and information (including advertising and sponsored content) that:

> is of a public or semi-public nature

> is shared or distributed via a digital platform

> has the potential to cause harm to an individual, social group or the broader community.

For the purposes of this paper, semi-public news and information refers to content that has been widely shared across a large but mostly insular online community or group, such as a Facebook group or WhatsApp group message. The ACMA considers this content should be reflected in the code given the growing role of messaging services and private online communities in spreading and amplifying misinformation.

Similarly, the ACMA considers that paid false, misleading or deceptive content should be in scope as it is promoted and distributed to a wide audience by platforms via their advertising services. These services can be used to specifically target individuals or groups that may be susceptible to its message, increasing potential harms. The control that platforms have over their advertising services, and the monetary benefit they obtain, place a greater responsibility on platforms to address the use of advertising services for the dissemination of misinformation.

To promote the implementation of measures to assist users to discern the quality of news and information on digital platforms, the ACMA considers that the code should also include some consideration of the criteria used in the selection, moderation or distribution of quality news. In particular, the code should encourage signatories to be transparent about the factors they take into account in assessing or signalling the quality of sources, including how these factors affect the visibility, discoverability and accessibility of quality news and information. This includes the roles played by algorithms as well as human moderators.

> **Position 2:** The ACMA expects the code to address misinformation across all types of news and information (including advertising and sponsored content) that is of a public or semi-public nature, distributed via digital platforms, and has the potential to cause harm. It also expects the code to cover platforms' considerations of what constitutes quality sources of news and information, and how this is communicated to users.

Code signatories

The code should, at a minimum, cover all digital platforms that were examined by the ACCC in the DPI, including online search engines, social media platforms and other major digital content aggregation services with at least one million active monthly Australian users. This is likely to capture the most widely used platforms, including Facebook, YouTube, Twitter, Google Search, Instagram, Snapchat, TikTok, LinkedIn, Google News and Apple News.

Of these, misinformation appears most prevalent on social media sites compared to search engines or news aggregators. Subsequently, platforms could consider whether there is a need for the code to stipulate different actions or activities based on these different business models.

Noting the continually evolving nature of misinformation, the ACMA considers the code could also apply to a range of other online services that play a growing role in distributing news and information to Australians. This could include, for example, virtual assistants and smart home devices (e.g. Amazon Alexa), online forums and other internet communities (e.g. Reddit), podcast aggregators (e.g. Spotify) and closed group messaging services (e.g. WhatsApp).

Most media organisations in Australia are already bound by a series of self and co-regulatory requirements for journalistic ethics and accuracy in news and reporting. As individual news outlets are not considered key distributors of misinformation in Australia, the ACMA does not consider they should be covered by the code.

Given the voluntary nature of this process, it will be a matter for individual platforms to decide on whether they participate in the development of the code or choose to be bound by the code. The ACMA would, however, strongly encourage all digital platforms with a presence in Australia, regardless of their size, to sign up to an industry-wide code to demonstrate their commitment to addressing misinformation. The ACMA's report to government in June 2021 will consider which platforms have signed up to the code. The government will take account of this report in assessing the need for further regulatory responses.

Changes in the industry within Australia (including new entrants to the market and changes in user bases) may lead to subsequent changes in code signatories. In administrating the code, participating platforms should monitor changes in the market and seek to encourage new signatories over time.

> **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.

Design of the code
Adopting an outcomes-based code and reporting framework

An outcomes-based regulatory approach involves establishing an agreed set of clear and measurable outcomes that describe what the code is seeking to achieve, while providing individual entities with the flexibility to decide how they deliver against these outcomes. An overview of outcomes-based regulation is provided below.

Adopting an outcomes-based regulatory approach

Outcomes-based regulation focuses on describing the outcomes or objectives that are to be achieved, without prescribing the means of doing so. This is in direct contrast to more traditional rules-based approaches, in which the same 'one-size-fits-all' model applies to all regulated entities in a market.

The primary distinguishing features of an outcomes-based approach are:

1. Regulation is drafted as high-level outcomes or objectives that must be met.
2. Entities develop their own systems to achieve the outcomes specified in the regulation.
3. 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.

Outcomes-based regulatory frameworks are particularly well suited to complex, dynamic and fragmented markets, where more traditional rules-based regulation is less able to keep pace with the rate of technological change. One such example is in the use of algorithms and machine learning tools to automate business processes. While regulators may struggle to establish detailed technical rules that stipulate how these systems are to work, under an 'outcomes' approach entities have the flexibility to develop their own innovative solutions to the problem in a way that best fits with their business. This approach also encourages greater levels of collaboration between regulators and industry, with both parties gaining a more detailed understanding of each other's activities and processes.

For outcomes-based regulation to be successful, there needs to be a level of cooperation and support by market participants, clearly defined and measurable outcomes, and a consistent program of evaluation and review to demonstrate compliance. If these conditions are not met, concerns may be raised about the efficacy of the underlying regulatory framework, as recently highlighted by the EU Code of Practice for Disinformation. As highlighted in chapter 3, reviewers have noted a lack of binding performance indicators in the EU Code and the insufficient provision of data to enable robust assessment of the EU Code's effectiveness.

Due to the rapidly changing market environment, the ACMA considers an outcomes-based approach would be the most suitable model for a voluntary code for addressing misinformation and news quality on digital platforms. It would expect platforms will have learnt from the EU code experience by including meaningful commitments to achieve outcomes that are carefully defined and clearly linked to the objectives of the code. As further discussed below, measures taken by platforms under this code must be capable of being monitored and measured over time, supported by a robust and transparent reporting framework.

Performance reporting is a key aspect of effective self-regulation and is especially important when considering an outcomes-based approach. While signatories would have considerable flexibility to determine what measures they adopt to meet the stated outcomes under the code, this comes with the responsibility of providing code administrators and reviewers with sufficient levels of information and data needed to demonstrate compliance and assess performance. Some examples of the types of information and data that could be provided by platforms is provided in Chapter 5.

This approach also recognises that platforms have a range of existing business models and different measures for addressing misinformation. It may not be appropriate to assess each platform's performance against a single or uniform set of industry-wide performance metrics.

Nevertheless, the ACMA expects the code to include a robust reporting regime that requires all signatories to publish an individual action plan. This plan should detail the various activities and initiatives to be undertaken by the platform to meet the outcomes of the code, key performance indicators that make it clear how the platforms will measure success, and any other ways in which the platform is responding to the evolving information environment. These plans should be finalised and made available to the public shortly after platforms have signed up to the code.

Signatories should also be required to report at least annually on their performance against their action plan, allowing for public scrutiny of the efficacy of the measures they have implemented. This reporting should include relevant and meaningful data to allow the ACMA or other organisations to conduct an independent, systematic and rigorous assessment of platforms' performance against the various provisions of the code. Platforms should also have their annual performance reporting independently audited, in line with best practice governance procedures.

> **Position 4:** The ACMA encourages platforms to consider an outcomes-based regulatory approach when developing their code. This should be supported by a strong performance reporting regime, requiring signatories to regularly publish performance indicators and report on their progress.

Activities to increase knowledge and capability

Platforms play an important role in building knowledge, capability and capacity to address online misinformation in Australia. Platforms should consider including specific commitments to research and associated activities in this context as part of the code development process.

Example commitments could include assisting and participating in research efforts to understand and combat information disorder, undertaking education campaigns and training programs relating to media literacy, and assisting efforts to increase the supply and quality of public-interest journalism and information where possible and appropriate. Platforms may also wish to formalise engagement with the education and research communities, including through the establishment of an advisory group and providing access to data to inform research efforts.

> **Position 5:** The ACMA expects the code will commit signatories to facilitate research, share relevant data, and undertake associated activities to improve understanding of misinformation in Australia. Platforms should consider ongoing avenues of collaboration between signatories, government, academia and other experts, and other relevant industries.

Code administration

Process to develop the code

In drafting and developing an industry-wide code, the ACMA would encourage all interested platforms and industry representatives to consider establishing or working through a dedicated representative body to co-ordinate input and progress this work. This would provide an opportunity for smaller or emerging platforms to participate equally in the process, while building on the collective regulatory experience of larger platforms that may have participated in other similar code development processes.

Prior to finalising the code, the ACMA expects that platforms will conduct a meaningful public consultation process. Input should be sought from experts across academia, relevant government agencies, and impacted stakeholders including consumer groups and users of the platform. Where possible, submissions and stakeholder feedback should be made publicly available.

In accordance with the government's implementation roadmap, the ACMA expects that a code will be in place by no later than December 2020. The ACMA will then report to government by June 2021 on the efficacy of the code.

> **Position 6:** The ACMA expects platforms to undertake an open, public consultation process when developing the code, with the code to be in place by no later than December 2020.

Administering and reviewing the code

The ACMA expects the code to require all signatories to implement a robust, effective and accessible complaints handling regime to address complaints about misinformation and platforms' performance under the code. Signatories should first try to address complaints internally through a transparent, responsive and user-friendly process. If they are unable to resolve the complaint internally, signatories should provide alternative dispute resolution (ADR), 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 designing and piloting an external dispute resolution scheme as part of its response to the DPI. Platforms may wish to consider whether this is an appropriate forum to address escalated complaints under the code.

The code should contain details on how it is to be administered, including through the formalisation of a representative committee or other body tasked to oversee its administration. This body should regularly convene to review the actions of signatories and monitor how they are meeting their commitments under the code. This body could also act as a decision-making forum, reviewing complaints made against the code, decide on actions for non-compliance by signatories, making amendments to the code and acting as an avenue for the sharing of new technologies, data, factchecks and initiatives between signatories.

The code should also include a statement about how it will be reviewed. This should include a mandated review after an initial time period (for example, 12 months or two years), as well as any key events or dates that could trigger the need for a subsequent review. The code should also detail a process to appoint an independent third-party reviewer, and outline the reviewer's functions and powers. This would help the code remain fit-for-purpose, providing an opportunity for platforms to make any necessary changes and address any real or perceived weaknesses.

> **Position 7:** The ACMA expects the code will require a robust, effective and accessible complaints handling regime. Users of digital platforms should also have access, free of charge, to an alternate dispute resolution process.
>
> **Position 8:** The ACMA expects a representative body will be established to oversee the administration of the code. The code should also include a mandatory code review mechanism.
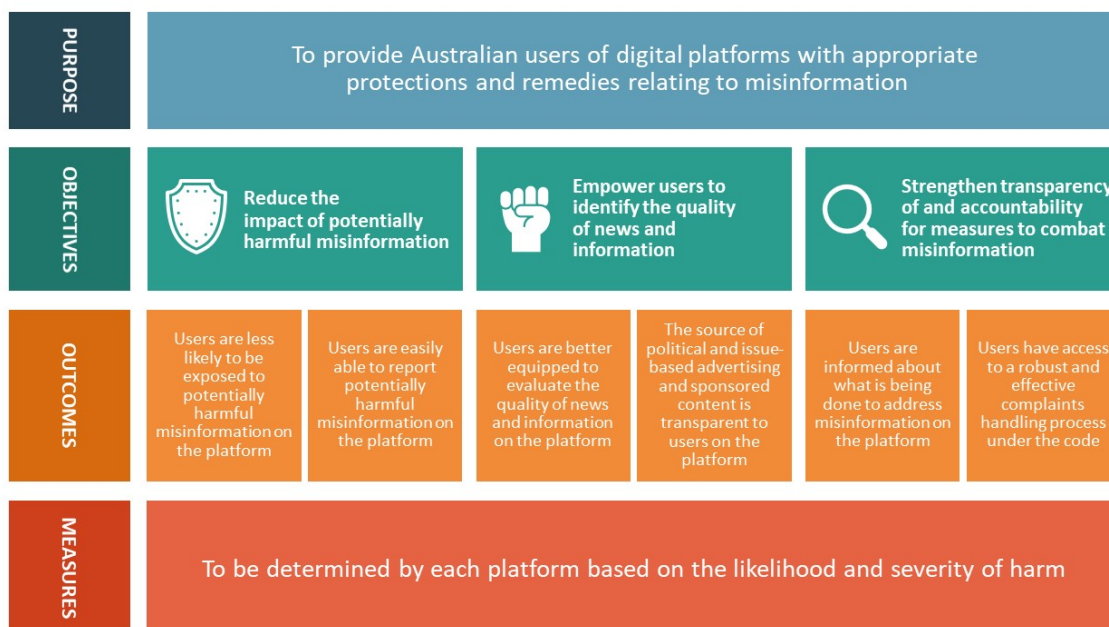
5. Proposed code model

As outlined in the previous chapter, the ACMA considers an outcomes-based regulatory approach would be the most suitable model for a voluntary code on misinformation and news quality.

To assist platforms, the ACMA has developed a proposed model for the code (Figure 2). This model sets out:

> the ACMA's expectation of what the code should achieve
> the potential objectives for the code
> the potential outcomes for the code.

This chapter will provide further detail on each layer of the model.

Figure 2:   Proposed outcomes-based model

PURPOSE	To provide Australian users of digital platforms with appropriate protections and remedies relating to misinformation				
OBJECTIVES	**Reduce the impact of potentially harmful misinformation**		**Empower users to identify the quality of news and information**		**Strengthen transparency of and accountability for measures to combat misinformation**
OUTCOMES	Users are less likely to be exposed to potentially harmful misinformation on the platform	Users are easily able to report potentially harmful misinformation on the platform	Users are better equipped to evaluate the quality of news and information on the platform	The source of political and issue-based advertising and sponsored content is transparent to users on the platform	Users are informed about what is being done to address misinformation on the platform	Users have access to a robust and effective complaints handling process under the code
MEASURES	To be determined by each platform based on the likelihood and severity of harm					

Purpose of the code

Preamble

The code should be prefaced with an opening, contextual statement that enables readers to understand why the code was established and what it seeks to achieve. In drafting this section, platforms should seek to define the problem of misinformation, identify the role of platforms in finding a solution, and commit signatories to take appropriate action.

Despite growing awareness and concern about the potential harms caused by misinformation, digital platforms provide an important avenue for the open exchange of news, information, opinion and expression across the Australian community. A preamble could provide platforms an opportunity to highlight this role, along with the need to carefully balance any additional scrutiny or oversight of user content with the rights of its users to freedoms of expression and privacy.

Where appropriate, the preamble could also acknowledge that platforms may already be subject to existing laws or regulatory arrangements that could overlap with some the matters covered by the code, and that in this context, those regulations will have

primacy. This could include, for example, the prohibition on false or misleading advertising under the Australian Consumer Law, advertising codes of conduct or defamation laws that protect individuals against false public claims resulting in reputational damage.

Definitions

Given the complexity of the subject matter and lack of consistent terminology in this space, platforms should include a standalone glossary or section on definitions at the beginning of the code. This would provide an opportunity to explain key concepts and assist the understanding of stakeholders and users.

Objectives of the code

A clear and unambiguous articulation of objectives—either as part of the preamble or on a standalone basis—can help stakeholders better understand the purpose and context of the code. Based on the findings of the DPI final report, the ACMA considers there are three overarching objectives that the code should seek to achieve.

Objective 1:	**Reduce the impact of potentially harmful misinformation** The code should seek to protect users of the platform and the broader community from harms caused by misinformation distributed via platforms.
Objective 2:	**Empower users to identify the quality of news and information** The code should seek to empower users to more easily detect misinformation on the platform and identify reliable and credible sources of news and information.
Objective 3:	**Strengthen the transparency of, and accountability for, measures to combat misinformation** The code should be based on best practice, with a focus on the principles of transparency and accountability.

While platforms may choose to adopt these specific objectives, the ACMA would recommend drafters consider, at a minimum, how the code will address the two separate but related arms of 'misinformation' and 'news quality'.

Outcomes of the code

One of the most important aspects of a code is establishing the set of rules or actions that signatories commit to undertake.

Under an outcomes-based model, drafters agree to a series of clearly defined and measurable outcomes, which should tie back to the overarching objectives of the code. Individual signatories are then responsible for implementing measures that meet these outcomes and for demonstrating compliance.

Reducing impact of misinformation

The ACMA considers a key objective of the code is to reduce the impact of, and exposure of Australian users to, potentially harmful misinformation on digital platforms. The table below provides examples of outcomes that address this broader objective and identifies a variety of means in which these outcomes could be met.

Objective 1: Reduce the impact of potentially harmful misinformation

Outcome 1: Users are less likely to be exposed to potentially harmful misinformation on the platform

> Platforms should have procedures, systems, and technologies in place to proactively identify and address misinformation, as well as accessible tools to enable users to identify and easily report instances of misinformation.

> Platforms should place their highest priority on limiting exposure to misinformation with a reasonable likelihood of:

> > causing significant personal injury or death

> > causing large-scale public panic or damage to public or private property

> > causing immediate and significant financial or economic harm

> > significantly undermining the integrity of elections or other major democratic processes in Australia.

> Platforms should adopt a harms-based approach when implementing measures to address misinformation. Severe and acute potential harms should be dealt with by stronger and more immediate actions that substantially reduce the impact or risk of exposure.

Examples of means through which this outcome could be implemented:

Use of detection algorithms; independent fact-checking services; proactive human monitoring and moderation; flagging, significant demotion or removal of offending content; notifying users sharing offending content; removal of malicious accounts.

Outcome 2: Users are easily able to report or flag potentially harmful misinformation on the platform

> Reporting tools or procedures should be easy to use and accessible, with multiple ways for users to contact the platform. Information on how to report harmful content should be promoted to users, with clear instructions on using reporting tools.

> Users should receive a notification that their report has been received and actioned, and as far as practicable, be notified of the status of the report and/or the result in a timely manner.

Examples of means through which this outcome could be implemented:

On-platform action menus and other off-platform reporting tools; clear and accessible instructions to flag or report misinformation; user notifications.

Empowering users to identify the quality of news and information

Beyond direct measures to reduce the impact of potentially harmful misinformation, the ACMA expects the code will also seek to address issues of news quality and credibility signalling. In particular, the ACMA's expectation is that platforms will provide adequate information and tools to help users:

> more easily detect and avoid misinformation

> more easily identify advertising and other forms of sponsored content that could otherwise be mistaken for user-generated content (particularly regarding political and issue-based advertising)

> identify accurate, reliable and timely sources of news and information, as discussed in Chapter 2.

Issue-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. These types of advertisements can be particularly dividing and sway public opinion on important matters. The ACMA considers it is important that both political and issue-based advertising adheres to electoral law obligations and is made transparent to the user (including the source of the advertisement) when this content is monetised on the platform.

Objective 2: Empower users to identify the quality of news and information

Outcome 3: Users are better equipped to evaluate the quality of news and information on the platform
> Platforms should implement measures and provide tools to help users identify quality and trustworthy news and information on the platform. Tools should be easy to use, and platforms should provide guidance or instruction where necessary for their use.   > Platforms should seek to enhance access to, and the discoverability of, high-quality news and information.   ***Example of means through which this outcome could be implemented:***   *Partnerships with independent fact-checking organisations; flagging suspicious content through content moderation tools; credibility signalling through flagging trusted sources and providing users with access to information on publishers and other sources; demoting misinformation; off-platform measures such as uplift of users' media literacy through campaigns/educational programs.*

Outcome 4: The source of political and issue-based advertising and sponsored content is transparent to users of the platform
> Platforms should clearly signal advertising and sponsored content and have processes in place to proactively identify misinformation in advertising or sponsored content before it is approved.   > The source of the advertisement and identity of the advertiser should be clearly distinguishable, and users should be able to readily access information on the advertiser and why their account is being targeted for such advertisements.   ***Examples of means through which this outcome could be implemented:***   *Clear labelling of advertisements and sponsored content; use of formatting to clearly distinguish ads from news articles and other information; clear disclosure of sponsored content; inclusion of links to policies on advertising and more contextual information about the advertiser, measures to pre-assess advertising or sponsored content.*

Measures to enhance transparency and accountability

In developing a code, platforms should look to review and adopt best-practice guidelines for industry codes and other forms of self-regulation and build on lessons from the EU code and other international responses.

To provide consistency across the measures implemented by platforms to address misinformation and news quality, the ACMA considers that the code should address the following principles:

> **Transparency:** ensuring government, industry and users have visibility of the code development process, the content of the code, and ongoing operation and effectiveness of the code over time, including what measures are being implemented to address misinformation and news quality.

> **Accountability:** ensuring code signatories are meeting their commitments through standardised complaints-handling processes, regular performance reporting and code review points, and agreed approaches for dealing with non-compliance (including details on who would review, enforce and oversee them).

Efforts to strengthen transparency and accountability obligations through the code should improve public trust in the quality of information available on digital platforms.

Objective 3: Strengthen transparency of and accountability for measures to combat misinformation

Outcome 5: Users are informed about what is being done to address misinformation on the platform

> Platforms should have clear information setting out how they are addressing misinformation. Measures should be applied in a consistent and transparent way and adjusted as necessary to respond to changes in the information environment.

> Platforms should publish their community guidelines, policies and procedures on misinformation in an accessible, user-friendly format in plain language. At a minimum, the documents should:

>   > Clearly outline the responsibilities of the platform and its users for the publication or distribution of information and

>   > Clearly state what actions the platform will take to address misinformation on their service, in proportion to the risk of harm, including specific measures for advertising and sponsored content.

> Platforms should widely promote this information, and actively inform users on- and off-platform when policies and procedures are updated.

   Examples of means through which this outcome could be implemented:

> *Plain language guides; informational videos; dedicated landing page on platform; tutorials; emails and user notifications.*

Outcome 6: Users have access to a robust and effective complaints handling process under the code

> Users should have an easily accessible avenue to submit complaints about the performance of the platform against the code and in its application of its published policies and procedures.

> Users should be notified of the status or outcome of their complaint within agreed timeframes and should have recourse to access independent alternative dispute resolution processes where they remain unsatisfied with the platform's response, at no cost.

> Platforms should report on the nature, volume and outcomes of complaints made under the code, and allow for independent oversight of the complaints process by any nominated external dispute resolution bodies or code reviewer(s).

Examples of means through which this outcome could be implemented:

Published complaints handling processes, accessible mechanisms to lodge complaints, independent body to review escalated complaints.

Measures under the code

Under an outcomes-based approach, the code itself should not be overly prescriptive or stipulate what actions or measures signatories are required to adopt in order to meet their obligations. Nevertheless, the code could provide some guidance as to what types of measures may be appropriate in different circumstances. To illustrate how this could work in practice under the ACMA's proposed regulatory model, it has provided some additional considerations which may assist platforms in developing their measures.

Harms-based approach

As discussed in Chapter 2, misinformation that is spread via digital platforms can result in a range of acute and chronic harms, with the likelihood and severity of these harms differing on a case-by-case basis, depending on a series of contextual factors. When assessing the risk of harm, platforms should be transparent about how they define harm, what factors they take into account as part of this assessment, and how they determine what level of harm is necessary to take action against offending content.

As with the risk of harm, measures to address misinformation and news quality concerns occur along a continuum. When designing misinformation measures, platforms should consider this continuum and adopt graduated, proportionate responses that increase in strength with the increasing risks of harm. Some examples of appropriate responses are outlined in the figure below, noting these will not be applicable to each platform or capture the full range of possible measures that could be undertaken to address a specific outcome.

Figure 3: Examples of measures based on risk of harm
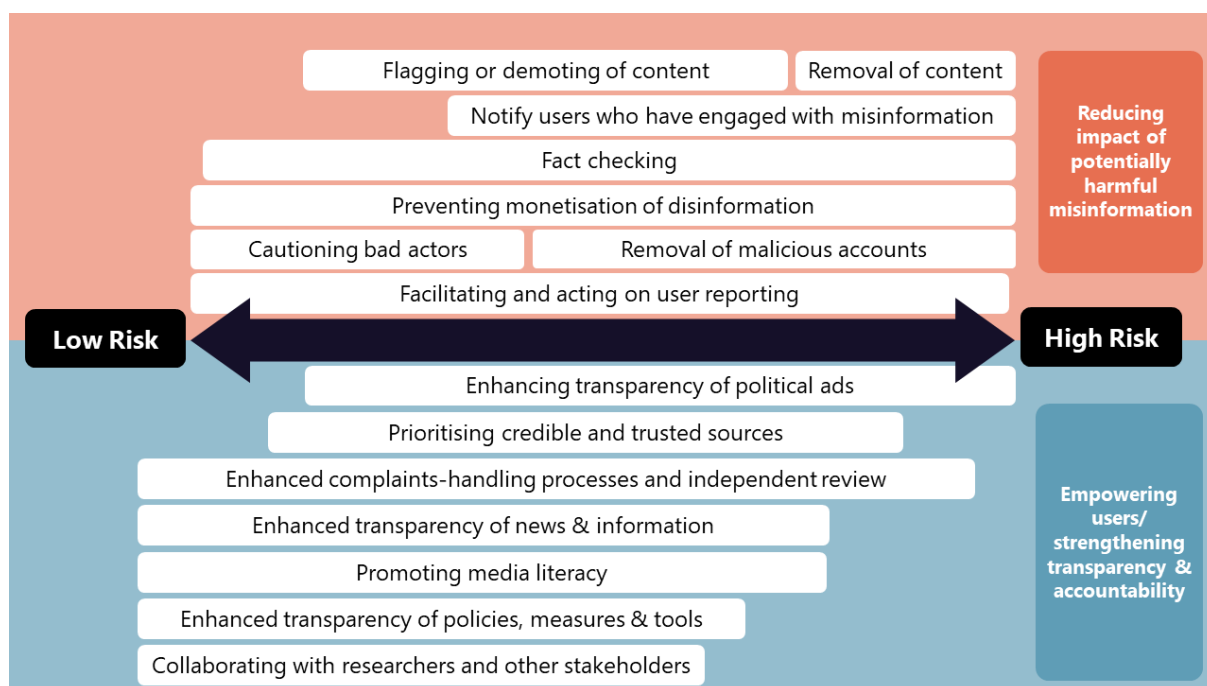
Content posing a high likelihood of acute and severe harm, like false medical advice that could result in a user's death, warrants an immediate and strong response from platforms. This could involve the significant demotion or removal of offending content and the removal of the offending user account. By contrast, measures designed to empower users to make more informed decisions about news quality would fall at the lower end of the continuum, and should involve less interventionist actions like the promotion of media literacy or enhancing transparency over how the platform determines quality.

Most platforms already have a series of graduated measures or sanctions in place for offending content. These are typically based on the severity of potential harm and built into community standards or terms of service. However, formalising a harms-based approach to measures under the code, including the provision of industry-wide guidance for assessing risk, would promote consistency across the industry and future proof the code by allowing platforms to adjust their measures in response to new developments.

Performance reporting

While the ACMA anticipates that the information reported by platforms will differ on a platform-by-platform basis, it expects the code to commit all signatories to develop key performance indicators and a reporting scheme that is based on the desired regulatory outcomes. Some suggestions about the kinds of information and data that could be provided by platforms include:

	Objective 1: Reduce the impact of potentially harmful misinformation
	> Information on the steps taken by the platform to make Australian users aware of how to report instances of harmful misinformation.
	> Data on user reports or user-flagged instances of potentially harmful misinformation targeting Australians on the platform.
	> Data on human-moderated or algorithmically identified instances of potentially harmful misinformation targeting Australians on the platform.
	> Data on the actions taken by the platform to address potentially harmful misinformation once identified (broken down by category such as take-downs and credibility signalling).
	> Data on the effectiveness of these actions and measures in reducing the volume and impact of potentially harmful misinformation on their platform.
	Objective 2: Empower users to identify the quality of news and information
	> Data and information about the measures the platform has in place to allow its users in Australia to readily discern the quality of news and information they access on the platform, such as credibility signalling.
	> Data and information about the measures the platform has in place to support media and information literacy among its users in Australia, both on and off the platform.
	> Information about the measures the platform has in place to allow its users in Australia to readily discern advertising and sponsored content on the platform.
	> Data to track the effectiveness of measures outlined in their action plan.
	Objective 3: Strengthen transparency of and accountability for measures to combat misinformation
	> Information about measures under development to address evolving forms of misinformation on the platform.
	> Data on complaints made about a platform's non-compliance with requirements of the code, and any dispute resolution processes and outcomes.

Platforms may choose to work constructively with the ACMA in determining their key performance indicators. The ACMA will be publishing further guidance on how it will measure the effectiveness of the code, and the performance of individual signatories, as part of its report to government in June 2021.

6. ACMA assessment

The ACMA has been requested by the government to oversee the code development process and to prepare a report on the adequacy of digital platforms' measures and the impact of disinformation more generally. Subject to further decisions by the government, this could be a once-off examination or part of an ongoing monitoring role for the ACMA.

In order to provide greater transparency to industry about this process, this section outlines the ACMA's current thinking on how to examine the effectiveness of the code—and other industry initiatives—in addressing misinformation and news quality issues in Australia.

Scope of the assessment

What activities are to be undertaken by the ACMA?

Although the ACMA is still in the preliminary stages of scoping its June 2021 report to government, the ACMA currently plans to focus on three separate areas of examination:

1. Assessing the code development process

While the ACMA does not anticipate playing any formal role in the drafting, administration or enforcement of the self-regulatory scheme, the ACMA will be closely overseeing and scrutinising the code development process. The ACMA's assessment will be based on observations about the strength and rigour of the process, including:

> the scale and breadth of public consultation

> the level of collaboration with government and other interested stakeholders

> the extent of cross-industry co-operation, including the number of signatories to one or multiple industry codes of practice.

Should there be no code in place by the time of the review, or if some major platforms choose not to participate in the process or become signatories, this may be regarded as evidence that a self-regulatory scheme is unlikely to be effective in addressing misinformation in Australia. Under these circumstances, the government could decide to develop a mandatory code or standard.

2. Assessing the content of the code(s) and resulting measures

The second area of examination will focus on the content and preliminary impact of the code or codes. The assessment will be based on whether the code contains all the necessary criteria for effective self-regulation, and the extent to which any new measures committed to by signatories can reasonably be expected to deliver on the desired regulatory outcomes outlined in Chapter 5.

To the greatest extent possible, this assessment will include an analysis of the strength and efficacy of any new measures being undertaken by signatories. Due to the limited timeframe between the expected finalisation of the code in December 2020 and the ACMA's report to the government in June 2021, the ACMA anticipates this initial assessment may need to rely on examining the effectiveness of existing activities, the reasonableness of any proposed measures as outlined in platforms' individual action plans, and point-in-time performance data detailed in progress updates. There may also be limited data available to the ACMA to make an assessment in some areas.

3. Assessing the state of online misinformation in Australia

The third proposed area for assessment is in response to the government's request that the ACMA's report include an examination of the broader impacts of disinformation in Australia. While the ACMA is still considering the full scope of this assessment, it considers it would be appropriate for the assessment to include both disinformation and misinformation.

Unlike the first two components, this would be an industry-wide view of the role of disinformation across the online information ecosystem in Australia. It may include the examination of specific case studies (including a retrospective assessment of 2019–20 bushfire season and the COVID-19 pandemic). There may also be an ability to identify a series of appropriate cross-industry metrics to baseline the state of misinformation and its impact on Australians. Should there be an ongoing monitoring role for the ACMA, this data would allow us to track changes to the state of misinformation over time.

Given the voluntary nature of the code and the possibility that some major platforms may choose not to participate, this third component of the ACMA's assessment will not be limited to code signatories. Should the ACMA identify any major non-signatory platforms of particular concern—either due to the size of their user base or the volume of identified misinformation—it may separately examine their role in the spread of misinformation in Australia and what, if any, measures the platforms have taken to address these concerns.

Sources of data and information

In order to undertake an assessment, the ACMA will need to source a variety of data and information on the code process, platforms' measures, and the impact of misinformation in Australia more generally.

Obtained from platforms

As outlined in Chapter 4, the ACMA expects the code will require signatories to publish individual action plans that outline how they intend to meet their obligations, including what data and metrics they will provide to track their performance. Platforms will then be required to report annually on their performance against these plans.

These action plans will be a valuable input into the ACMA's assessment, providing an overview of the measures that platforms are planning to implement to address misinformation. The ACMA would also request that all signatories provide a written progress report to the ACMA by no later than April 2021, outlining how they are working to meet their commitments under the code. This progress report will be a critical input into the ACMA's 2021 assessment. While this initial report may not be as detailed as subsequent annual reports, the ACMA anticipates that the reporting obligations of signatories would continue to evolve and become more detailed over the longer term in line with new reporting procedures and systems, greater levels of industry collaboration and the changing market.

The ACMA will also be seeking to meet with interested platforms on a semi-regular basis to discuss what progress has been made towards the code development, and encouraging all major platforms to work with us over the coming months to identify what data and information that the ACMA could rely on to inform its baseline impact assessment.

Other sources

There are a range of additional or supplementary sources of information and data that the ACMA is currently exploring to assist in its assessment, and that may form part of the broader reporting framework. This could include, for example, qualitative and quantitative consumer research to better understand how Australians perceive and engage with misinformation, and to examine the impact of specific measures.

Other forms of monitoring and analysis of misinformation, such as propagation mapping or sentiment analysis, could also be used to better understand the state and impact of misinformation on the Australia information ecosystem. This work would allow the ACMA to examine and report on specific case studies on the spread and general effects of misinformation on Australian society, such as on the COVID-19 pandemic as discussed further in Appendix A.

The ACMA may consider working with third-party research firms, or partnering with the Australian academic community, to assist in capturing and analysing data to inform its assessment.

Proposed process

To enable a transparent, fair and objective assessment, the ACMA is proposing to establish a 'misinformation and news quality monitoring and reporting framework'. This document will provide industry with further guidance on how the ACMA will set a baseline and measure the effectiveness of the code, as well as evaluate the overall impact of platforms' contributions to misinformation in Australia.

As part of the development of this framework, the ACMA will work in collaboration with platforms and experts across academia and industry on identifying key metrics and indicators. These metrics will serve dual purposes:

> to assist platforms in determining the most appropriate way to measure and report their performance against the outcomes under the code

> to help the ACMA in its broader assessment work of both individual platforms and cross-industry impacts of misinformation.

Following the workshops and publication of the monitoring framework, the ACMA may commission further research to support its assessment or request additional data from platforms.

Figure 4:   Proposed activities and process for the ACMA and industry
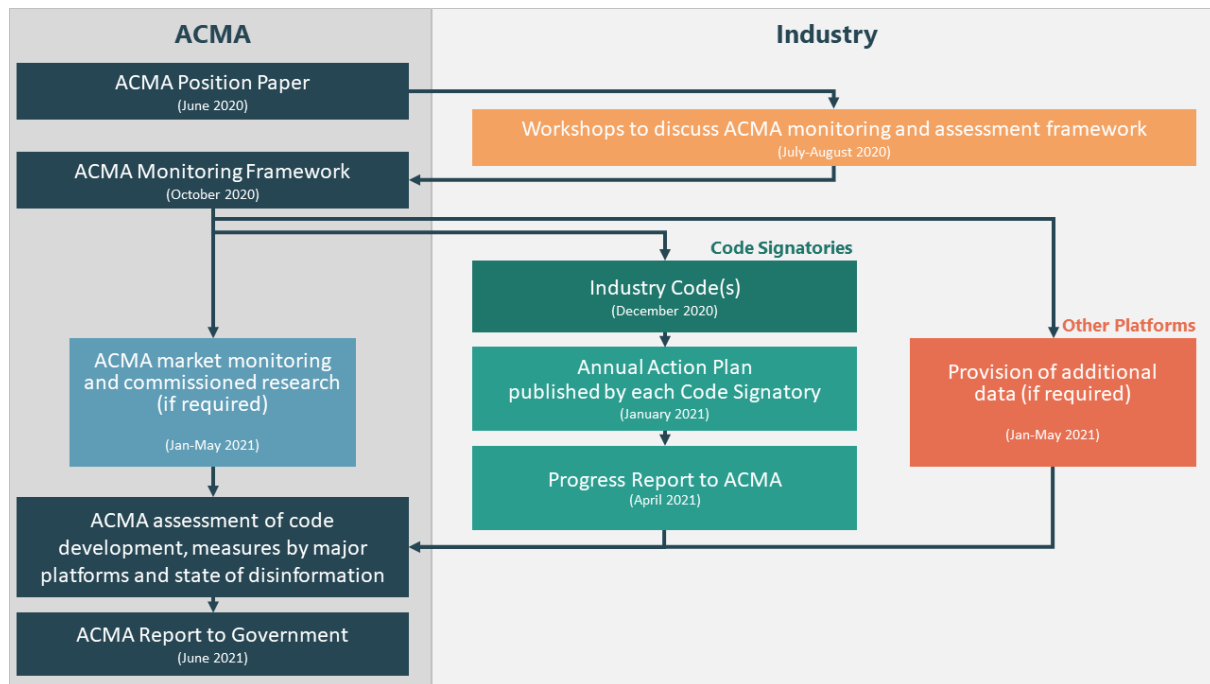
The ACMA's assessment and report to government in June 2021 will provide evidence on whether platforms have delivered an effective voluntary code. This report will help inform the government's consideration on next steps including on whether further reforms or regulatory responses are required. Should the ACMA determine that the voluntary code or codes are ineffective or fail to provide sufficient protections to users of digital platforms, the ACMA may recommend to the government that a regulatory framework applicable to digital platforms be put in place.

7. Next steps

The ACMA will continue to develop its monitoring and reporting framework in parallel with the platforms developing their voluntary code. The ACMA considers the reporting framework set out in the code will be key inputs.

To align both processes, the ACMA will hold a range of workshops with digital platforms, academia and other government agencies to inform the development of its monitoring and reporting framework. The ACMA anticipates holding these workshops early in the 2020-21 financial year.

The ACMA expects its framework will be finalised by October 2020. This timing provides an ability for the ACMA to consider a draft code released for public consultation by platforms. The ACMA will publish a summary of its framework to provide additional transparency about its activities.

In early 2021, the ACMA will transition to a purely monitoring role to inform its first report to government in June 2021. A proposed timeline of key activities and indicative dates is below.

Phase	Activity	Dates
Position paper	ACMA releases position paper	June 2020
Code development	Digital platforms develop code(s) including public consultation on draft code(s)	June 2020 to December 2020
Monitoring and assessment framework	ACMA holds workshops to help progress a monitoring and reporting framework	July 2020 to August 2020
	ACMA finalises monitoring and reporting framework	October 2020
Monitoring and implementation	Digital platforms and ACMA determine information and data sources to establish a robust baseline for monitoring and evaluation	May 2020 to January 2021
	ACMA public update on monitoring framework and voluntary codes	January/February 2021
	ACMA monitors environment for misinformation and news quality issues	June 2020 onwards
	ACMA and digital platforms discuss implementation of measures in accordance with the code(s)	January 2021 to June 2021
	Code signatories to provide update to ACMA to inform report to government	April 2021
	ACMA's first report due to government	June 2021

Appendix A: Case study— misinformation during COVID-19

The COVID-19 pandemic has been accompanied by such a high volume of potentially harmful misinformation that the World Health Organisation (WHO) has described it as an infodemic.[31]

The ACMA has considered COVID-19 as a case study to examine the types, sources, and content of misinformation, as well as the responses to it from digital platforms.

The novel nature of the virus, its rapid spread, and the severity of the public health measures required to control it, have together created an environment of great public uncertainty. Media and news consumption have risen markedly, especially online, with the public seeking reliable information on the virus and the latest news. High levels of concern about the virus outbreak are correlated with higher news consumption.[32] Both traditional media usage and social media usage have increased, especially during the early stages of the outbreak.[33]

This has created an ideal environment for the spread of misinformation. The sheer volume of misinformation seen during the crisis, combined with increased digital media usage, means a large proportion of users are potentially exposed to misinformation, and information from official sources of information can be drowned out.

Some of this misinformation has the potential to lead to serious harm to individual health, as well as exacerbating the public health crisis and undermining trust in civil institutions. Malicious actors are also taking advantage of the crisis to conduct disinformation campaigns and to lure users into scams.[34]

Volume of misinformation

Due to the ongoing nature of the crisis, the little research that has been published to date is preliminary, restricted in scope or has the potential to become quickly out of date. Data for many platforms is also not readily available. Despite this, available figures show cause for concern.

Using machine learning to analyse data from third-party fact-checkers, an Italian research group at the Bruno Kessler Foundation:

> Identified 5.9 million posts on Twitter from March linking to news on COVID-19. Of these, 1.7 million shared links to sites containing misinformation.[35]

> Found that engagement with misinformation appeared particularly strong in the early stages of the pandemic. A shift towards more reliable sources occurred as the contagion spread further.[36]

---

[31] World Health Organization, 'Munich Security Conference', 15 February 2020.

[32] S Park et al., COVID-19: Australian news and misinformation, News and Media Research Centre, University of Canberra, May 2020.

[33] ibid., p. 15; see also Ofcom's ongoing research into media consumption and attitudes during COVID-19.

[34] R Heilweil, Coronavirus scammers are flooding social media with fake cures and tests, *Vox*, 17 April, 2020.

[35] E Hollywood & A Mostrous, 'Fake news in the time of C-19' *Tortoise*, 23 March 2020.

[36] Fondazione Bruno Kessler, 'Covid-19 and fake news in the social media', 10 March 2020.

> Notes Twitter is only one platform, and its user base is smaller than Facebook, Google, or YouTube. The true volume of and exposure to misinformation across social media would be substantially higher.[37]

As of 6 April, the International Fact Checking Network reported having debunked over 3000 individual pieces of misinformation.[38]

In research released on 16 April, civil campaign group Avaaz analysed a sample of 104 Facebook posts containing COVID-19-related misinformation debunked by independent fact checkers. They found that the posts in the sample were shared over 1.7 million times on Facebook and viewed an estimated 117 million times. This does not include interactions with 'clones' of the sampled content.[39]

Reports on the sharing of *Plandemic*, a film suggesting that COVID-19 has been deliberately spread by a global conspiracy, state that in the first week after its release on social media it was viewed over eight million times, despite efforts by platforms to remove or demote it.[40]

While most research necessarily focuses on open platforms, there is also evidence that significant amounts of misinformation have been shared on private groups and platforms.[41]

Reports from digital platforms also provide an indication of the volume of misinformation circulating. Facebook has reported that during March, they placed warnings on about 40 million posts related to COVID-19, based on around 4,000 articles by their independent fact-checking partners.[42] It also reported removing hundreds of thousands of pieces of misinformation.

Types of misinformation
A report from the Reuters Institute describes the results of a systematic content analysis of misinformation from a corpus of fact-checked articles through to the end of March.[43]

The content analysis categorised the misinformation into five types according to the level of fabrication.[44]

37 E Hollywood & A Mostrous, '[Fake news in the time of C-19](#)' *Tortoise*, 23 March 2020.

38 E Oputa, '[How fact-checkers are staying ahead of the COVID-19 whack-a-mole](#)', *Poynter.*, 14 April 2020.

39 AVAAZ, *[How Facebook can Flatten the Curve of the Coronavirus Infodemic](#)*, 2020.

40 C Newton, '[How the 'Plandemic' video hoax went viral](#)', *The Verge*, 12 May 2020.

41 Institute for Strategic Dialogue, *[Covid-19 Disinformation Briefing No.1](#)*, 27 March 2020, p. 2 and *[COVID-19 Disinformation Briefing No. 2](#)*, 9 April 2020; A Ferrante, '[COVID-19: Attempts to influence and deceive](#)', The Hill, 20 April 2020.

42 G Rosen, [An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19](#), Facebook Newsroom, 16 April 2020.

43 JS Brennen et al., *[Types, Sources, and Claims of COVID-19 Misinformation](#)*, Reuters Institute for the Study of Journalism, University of Oxford, 2020.

44 The schema is adapted from Clare Wardle of First Draft. The latest version of Wardle's schema can be found in *[First Draft's Essential Guide to Understanding Information Disorder](#)*, First Draft, 2019. Wardle's schema has seven categories, with *false context* used to describe genuine information shared with false context. The Reuters Institute study classifies this as misleading content. Wardle uses *false connection* in the same sense as *false context* is used by Reuters.

> **Misleading content:** contains some true or genuine information, but the details have been selected, reformulated or recontextualised in a misleading way.

> **False context:** images or videos with misleading headlines, captions or descriptions.

> **Manipulated content:** contains manipulated images or videos.

> **Fabricated content:** information is completely fabricated.

> **Imposter content:** a genuine source is misrepresented as the author of the content.

Note: satire or parody content is identified separately.

The researchers found that the majority of the misinformation (59 per cent) comprised reconfigured content, with existing and often true information spun, twisted, recontextualised, or reworked. Only 38 per cent was completely fabricated.

The most common form of misinformation was misleading content (29 per cent), which contained some true information, but where details were reformulated, selected, and re-contextualised in ways that made them false or misleading. The prevalence of these different types of misinformation can be seen in the following figure.

Figure 5: Reconfigured vs fabricated misinformation
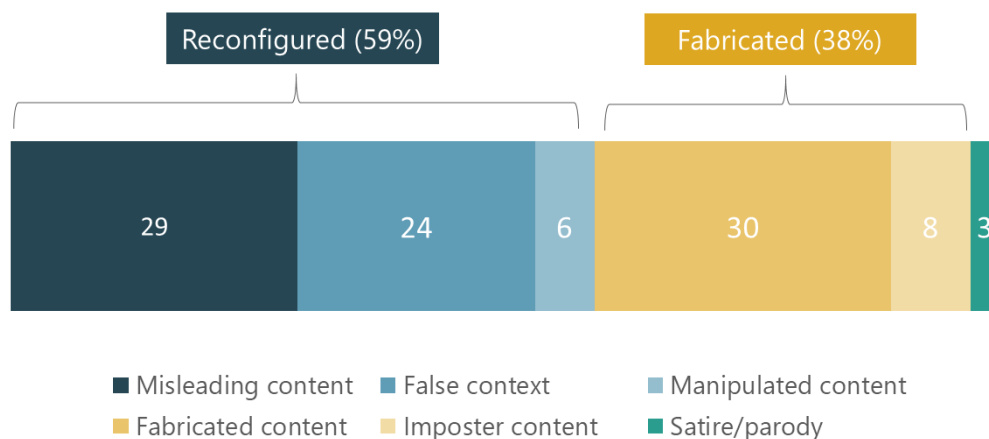
Note: Shows the proportion of reconfigured (n=133) and fabricated (n=86) misinformation in the sample (n=225) and the types of misinformation that constitute both.
Source: JS Brennen et al., Types, Sources, and Claims of COVID-19 Misinformation, Reuters Institute for the Study of Journalism, University of Oxford, Factsheet April 2020.

Significantly, users engaged much more extensively with reconfigured content (87 per cent of interactions) than wholly fabricated content (12 per cent).

Also noteworthy is the lack of sophisticated deep fakes: every example of doctored or manipulated content in the sample employed low-tech photo or video editing techniques. This suggests that sophisticated deception is not required to promote sharing amongst users. The ACMA notes, however, that the overwhelming volume of information and the emotive nature of COVID-19 mean this may not hold during normal circumstances.

Sources of misinformation

Top-down vs bottom-up

In the same Reuters Institute study, misinformation from politicians, celebrities, and other prominent public figures ('top-down' misinformation) made up just 20 per cent of the claims in the sample but accounted for 69 per cent of total social media engagement. Most of these are social media posts, though they include politicians speaking publicly or to the media.

While most individual pieces of misinformation on social media came from the public ('bottom-up misinformation'), on the whole these posts generated far less engagement (30 to 31 per cent) than top-down misinformation. Only a few instances of bottom-up misinformation achieved a large reach. The researchers point out that their analysis is limited to publicly availably posts on social media and does not account for misinformation spread in private groups and via messaging services, which other reports suggest are significant platforms for significant amounts of bottom-up misinformation.[45]

A study by NewsGuard on Facebook identified 31 pages as 'superspreaders' of COVID-19 misinformation. These pages had a combined reach of over 21 million followers.[46] NewsGuard also found that sites that spread COVID-19 misinformation are very likely to have spread other kinds of misinformation in the past. The engagement bias towards top-down misinformation likely indicates the influence of large cohorts of followers and may be increased by the algorithmic promotion of trending content.

Malicious vs non-malicious intent

The Reuters Institute study also classifies misinformation on the basis of apparent motivation but notes that motivation is difficult to assess from content alone. It found that the pandemic appears to be supplying an opportunity for very different actors with a range of different motivations and goals to spread a wide variety of misinformation.

Very few pieces of misinformation appeared to be shared for financial gain. The researchers note, however, that as the source of their sample is a fact-checking database, this likely excludes the majority of low-grade spam and scam material. The ACMA has received a large number of complaints relating to spam and scam emails and SMSs, which may reflect a significant volume of this material in circulation. The Bruno Kessler Foundation study found that 40 per cent of the tweets posting misinformation came from accounts controlled by bots.[47]

A great deal of misinformation appears to be created or promulgated by foreign actors, including state-affiliated actors engaged in disinformation campaigns.[48] Fringe and extremist groups in democratic countries are also engaged in spreading mis- and disinformation, especially conspiracy theories.[49]

Many of these messages are believed and spread further by unwitting users. This includes those who lack access to definitive, accurate information (including where

---

45   Institute for Strategic Dialogue, *Covid-19 Disinformation Briefing No.1*, 27 March 2020, p. 2.

46   Newsguard, Tracking Facebook's COVID-19 Misinformation 'Super-spreaders', 23 April 2020.

47   M de Domenico & P Sacco, *COVID-19 Infodemics: A Report on the Current State of (Dis)information*, 9 March 2020.

48   European External Action Service, EEAS Special Report Update: short assessment of narratives and disinformation around the covid-19 pandemic, 20 May 2020.

49   Institute for Strategic Dialogue, *Covid-19 Disinformation Briefing No.1*, 27 March 2020, p. 2 and *COVID-19 Disinformation Briefing No. 2*, 9 April 2020.

such information is not available or unclear) and those who are distrustful of information provided by government or other official sources.[50]

Content of misinformation

Although the content of COVID-19 misinformation is continually evolving, some studies have analysed the prevalence of certain topics at particular points in time.

The Reuters Institute study classifies COVID-19 misinformation from March 2020 according to the topic or nature of the claims being made. The largest category of false or misleading claims concerns the actions or policies of public authorities, including government and international bodies like the WHO or the UN, appearing in 39 per cent of the misinformation they analysed. The next most common categories comprise claims about the spread of the virus, general medical advice (including cures) or misinformation about prominent public figures.

Figure 6:   Percentage of sample containing claim related to different topics
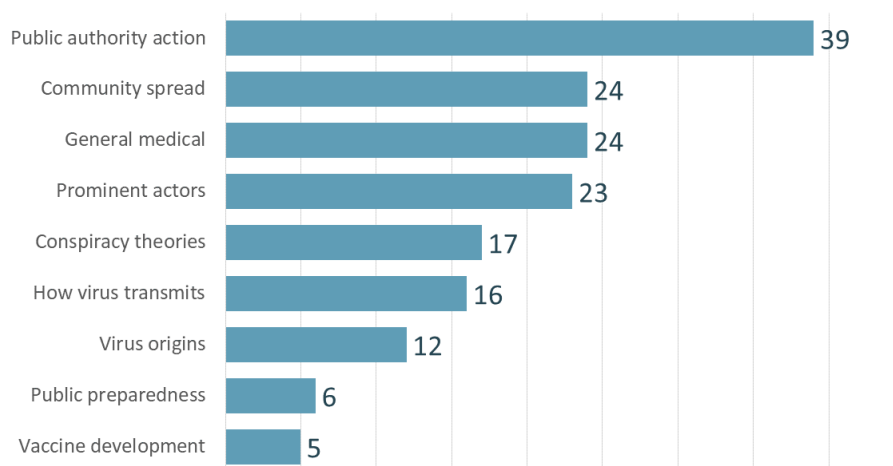
Note: Based on proportion of sample (n=225) containing each claim. Pieces of misinformation may contain multiple claims.
Source: JS Brennen et al., Types, Sources, and Claims of COVID-19 Misinformation, Reuters Institute for the Study of Journalism, University of Oxford, 2020.

The Bruno Kessler Foundation study found that the largest proportion of false virus-related news on Twitter in March related to fake events or occurrences.[51]

The International Fact Checking Network reports that misinformation has typically followed a distinct pattern in each country as the virus has spread, coming in distinct waves beginning with misinformation about the origins of the virus, moving to false medical information about cures or susceptibility of different groups and then to government and community responses.[52]

The Atlantic Council's Digital Forensic Research Lab also highlights the problem of outdated information in such a rapidly changing scenario.[53] Information that was accurate a week ago may not be accurate today. In many cases outdated information

---

50   JS Brennen et al., *Types, Sources, and Claims of COVID-19 Misinformation*, Reuters Institute for the Study of Journalism, University of Oxford, 2020.

51   E Hollywood & A Mostrous, 'Fake news in the time of C-19' *Tortoise*, 23 March 2020.

52   E Suarez, 'How fact-checkers are fighting coronavirus misinformation worldwide' Reuters Institute, 31 March 2020.

53   DFR Lab, 'When accurate coronavirus info grows stale, obsolete data becomes misinformation', 4 April 2020.

is shared innocently, but it has also been picked up and deliberately shared in active disinformation campaigns.

Some misinformation links to pre-existing conspiracy theories. Anti-vaccination groups have disseminated conspiracy theories about the origins of the virus,[54] and the notion that COVID-19 is related to the rollout of 5G has spread rapidly.

Originating from an interview with a Belgian conspiracy theorist, the theory spread to niche communities on closed and mainstream social networks, where it was shared further by fringe publications and disinformation agents. As the theory spread, it was picked up by celebrities and amplified by social media algorithms. It appeared on mainstream media as the theory translated into practical action, with conspiracy theorists launching attacks of vandalism on 5G transmitter sites in the United Kingdom.[55]

According to Ofcom research, 5G-related theories were the most common misinformation UK respondents encountered on social media in April and May 2020.[56] Protests about governments' COVID-19 measures have also occurred in the USA, Australia, and elsewhere, with many protestors focusing on purported links between the pandemic and 5G transmitters or vaccinations, as well as a range of other conspiracy theories.[57]

Impact of misinformation

The impact of misinformation is difficult to measure, particularly so close to the events in question. There have been reports of people following harmful advice relating to cures and preventive measures.[58] As noted earlier, the 5G conspiracy theory has led to damage to public infrastructure. Misinformation on government measures and the impact of the virus may also have contributed to social unrest.

Recent surveys also give insight into consumer sentiment. Research conducted between 18 and 22 April by the University of Canberra found that 23 per cent of Australians have encountered a lot of misinformation and a further 36 per cent some misinformation.[59] Social media is the most common source of misinformation, with 66 per cent of those having encountered misinformation reporting they encountered it on social media, compared to 36 per cent on news media and 30 per cent having been forwarded misinformation from someone they know.[60]

---

54   DFRLab, 'Bill Gates trends in South Africa after coronavirus anti-vaxxer conspiracy theory goes viral', 4 April 2020.

55   J Temperton, 'How the 5G coronavirus conspiracy theory tore through the internet', *Wired*, 6 April 2020.

56   Ofcom, Covid-19 news and information: consumption and attitudes, 2020.

57   J Taylor, 'Vaccines, 5G, Bill Gates: why are Australians gathering to spread coronavirus conspiracy theories?', *The Guardian,* 12 May 2020; M Nienaber & T Escritt, 'Merkel urges prudence as Germany's infection rate remains critical', *Reuters*, 11 May 2020; E Young, 'Why the far-right and militia groups are protesting coronavirus lockdowns in the US', *SBS News*, 23 April 2020.

58   T Waldrup et al., 'Fearing coronavirus, Arizona man dies after taking a form of chloroquine used to treat aquariums'.

59   S Park et al., *COVID-19: Australian news and misinformation*, News and Media Research Centre, University of Canberra, 2020, p. 26. The findings are reflective of Australians with access to the internet.

60   ibid. According to the report, those with higher education levels more frequently report having encountered misinformation, suggesting that higher media literacy may lead to higher detection rates.

Overall trust in COVID-19 news is higher than general trust in news media as reported in early 2020.[61] This includes trust in news on social media, although trust in social media is generally low and in a downward trend. Twenty-one per cent of Australians reported trusting news on social media during COVID-19, compared with 17 per cent generally (down from 18 per cent in 2019).[62]

Misinformation appears to drive news avoidance, with 79 per cent of those who have encountered misinformation about COVID-19 avoiding news compared to 69 per cent of those who have encountered misinformation about the virus a little or not at all. Sixteen per cent directly reported not knowing whether news is true as a reason for avoiding news on COVID-19.[63] Many report changes in their behaviour of the course of the pandemic, with rising consumption of and attention paid to traditional media and shrinking consumption of and attention paid to social media as the pandemic has progressed.[64]

Misinformation also appears to be influencing belief. The University of Canberra reports that 19 per cent of Australians thought it was true that the coronavirus was made in a lab, while 40 per cent said they didn't know.[65] The Essential Report found that 39 per cent of Australians believe this to be definitely or probably true. It also found that 20 per cent of Australians believe it is definitely or probably true that the number of COVID-19 deaths has been exaggerated by the media and governments to scare the population, 13 per cent believe that it is definitely or probably true that the COVID-19 virus is not dangerous and is being used to force people to get vaccines, and 12 per cent believe that the 5G wireless network is being used to spread the COVID-19 virus.[66] There is also evidence that engagement with related misinformation such as anti-vaccination posts has risen during the pandemic.[67]

EU researchers have noted that repeated exposure to particular false stories can increase the likelihood of belief in the story.[68]

In the UK, the first tranche of a weekly Ofcom consumer survey on COVID-19 found that over 50 per cent of those surveyed reported having encountered false or misleading information in the previous week. Seven per cent of those who saw misinformation reported having shared it. These figures dropped slightly over following weeks of the survey. In week 6 (May 1–3), 38 per cent said they found it hard to know what is true and what is false about the coronavirus, in line with results from week one (40 per cent).[69]

Official sources and traditional broadcasters remain the most trusted sources for information on COVID-19, with social media and closed groups continuing to be the least trusted. The proportion of respondents using a fact-checking site has increased over time, and the proportion of respondents who say they are 'finding it hard to know

61 ibid., p. 24. For comparison, see S Park, et al., *Digital News Report: Australia 2020*, News and Media Research Centre, University of Canberra, 2020, p. 73.

62 ibid., p. 24. For comparison see S Park, et al., *Digital News Report: Australia 2020*, News and Media Research Centre, University of Canberra, 2020, p. 74.

63 ibid., p. 19.

64 ibid., pp. 28–9; See also Ofcom, Covid-19 news and information: consumption and attitudes, 2020.

65 S Park et al., *COVID-19: Australian news and misinformation*, News and Media Research Centre, University of Canberra, 2020, p. 31.

66 Essential Research, *The Essential Report*, 18 May 2020, p. 10.

67 C Wilson, 'As The World Hopes For A COVID-19 Vaccine, Anti-Vaxxers Are Growing Their Social Media Influence', *Buzzfeed*, 20 May 2020.

68 Repeating a lie does not make it true', *Disinfo Review*, 9 April 2020.

69 Ofcom, Covid-19 news and information: consumption and attitudes, 2020.

what's true and what's false about Covid-19' has decreased from 40 per cent in week 1 to 32 per cent in week 3.

Responses from platforms

Policy changes

On 17 March, Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter and YouTube issued a joint statement outlining their approach to addressing misinformation about COVID-19 on their platforms.[70]

This appeared to mark a turning point for many platforms, indicating they would take a broader approach to countering misinformation beyond the realms of the coordinated inauthentic activity that often accompanies malicious disinformation campaigns. Concrete changes to platform policies include broadening rules on content review, removal and sharing.

> WhatsApp has announced measures to address the spread of misinformation on its messaging platform, limiting the number of recipients for frequently shared messages.

> Twitter has broadened its definition of harm to 'address content that goes directly against guidance on COVID-19 from authoritative sources of global and local public health information'. It will prioritise the removal of content when it has a call to action that could potentially cause harm. Twitter has also begun labelling tweets containing potentially harmful misinformation and providing links from these tweets to authoritative sources. Tweets confirmed as containing misinformation with a high risk of harm will be removed, including those from world leaders.[71]

> YouTube policies now state that 'any content that disputes the existence or transmission of COVID-19 … is in violation of YouTube policies' and will be removed.

> Facebook has also broadened its policies to allow for the removal of false claims which could lead to physical harm.[72]

> Google (including YouTube) updated its advertising policy to state that public health crises are covered by its sensitive events policy, which prohibits advertisements seeking to profit from such an event where there is no discernible benefit to users.[73] Google states it will also remove ads that discourage people from seeking medical treatment or claim harmful substances have health benefits.

Collaboration

Platforms have announced increased collaboration with governments, health authorities, experts and third-party fact-checkers to help counteract COVID-19 misinformation online.

> Facebook is allowing national ministries of health and reliable organisations to advertise accurate information on COVID-19 free of charge. Twitter, which prohibits political advertising, is allowing links to the Australian Department of Health and World Health Organization websites.

---

[70]   https://twitter.com/fbnewsroom/status/1239703497479614466

[71]   Y Roth & N Pickles, 'Updating our Approach to Misleading Information', *Twitter: blog*, 11 May 2020.

[72]   G Rosen, An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19, *Facebook Newsroom*, 16 April 2020.

[73]   Coronavirus disease (COVID-19) Google Ads policy updates, *Google Ads Help*, updated 30 March 2020.

> Rather than relying on user reports, Twitter is enforcing their expanded content-review policies in close coordination with trusted partners, including public health authorities and governments, and continue to use and consult with information from those sources when reviewing content.' Twitter is also working to give the tick of authenticity to health experts.

> Google is providing information from authoritative sources from its homepage, and searches relating to the COVID-19 pandemic will bring up an 'SOS alert' (and 'panels' on YouTube) with information from and links to authoritative sources.

Actions taken to address misinformation

Facebook has stated it has taken a range of actions, including:

> Expanded its network of fact-checkers in response to COVID-19 misinformation.[74] If a piece of content is rated false by fact-checkers, Facebook will work to reduce its distribution by demoting it in user feeds and show warning labels that link to contextual information. Users who have previously liked, shared or commented on a post later tagged as false are notified, and private groups that spread falsehood will no longer show up in recommendations to other users. Facebook's detection algorithms can also learn from fact-checks to identify duplicate misinformation. Facebook stated that when users saw fact-checking warning labels, 95 per cent of the time they did not go on to view the original content.

> Removing hundreds of thousands of pieces of misinformation that could lead to imminent physical harm, including claims that drinking bleach cures the virus and that physical distancing is ineffective in preventing the disease from spreading.

Facebook, YouTube and Twitter reported removing coronavirus misinformation posted by politicians and public figures, which historically they have been reluctant to do.

Conclusions

Above all, the magnitude of the COVID-19 pandemic has highlighted the complex and rapidly evolving nature of the problem of online misinformation.

The COVID-19 infodemic has also brought home that combating malicious behaviour from state actors and scammers is only one facet of misinformation, which is a far broader issue requiring a multi-pronged response.[75]

Platforms have taken substantial steps to combat misinformation, by updating policies, by collaborating with governments, other authorities and external partners and by acting on the vast amount of misinformation circulating on their platforms.

Recent data and research show that while these efforts have shown some success[76], misinformation continues to spread.[77] As argued by J Scott Brennen and his colleagues at the Reuters Institute, there will be no silver bullet or inoculation for misinformation about the new coronavirus.[78] Instead, addressing the spread of misinformation about COVID-19 will take a sustained and coordinated effort.

---

[74] G Rosen, 'An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19', *Facebook Newsroom*, 16 April 2020.

[75] G Machado, 'Being cautious with attribution: Foreign interference & COVID-19 disinformation' *EU Disinfo Lab*, 10 April 2020.

[76] See e.g., 'Coronavirus: Viral WhatsApp messages 'drop 70%'', *BBC News*, 27 April 2020.

[77] Based on recent research and data at time of publication. See e.g. Ofcom's ongoing Covid-19 news and information: consumption and attitudes research; Avaaz's Facebook study.

[78] JS Brennen et al., *Types, Sources, and Claims of COVID-19 Misinformation*, Reuters Institute for the Study of Journalism, University of Oxford, 2020.

Platforms, public authorities, independent fact-checkers and news media must come together to help the public understand and navigate the pandemic.

This crisis provides us with an opportunity to engage collaboratively to address the problem of online misinformation with a considered and principled approach that will help platforms, governments and users respond effectively to new crises of information.

Appendix B: Consolidated list of ACMA's positions

Position 1

The ACMA encourages platforms to consider a single, industry-wide code that provides appropriate protections and remedies for Australian users of digital platforms. It expects this code will be consumer centric, readily accessible to the public, and fit-for-purpose for Australia.

Position 2

The ACMA expects the code to address misinformation across all types of news and information (including advertising and sponsored content) that is of a public or semi-public nature, distributed via digital platforms, and has the potential to cause harm. It also expects the code to cover platforms' considerations of what constitutes quality sources of news and information, and how this is communicated to users.

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.

Position 4

The ACMA encourages platforms to consider an outcomes-based regulatory approach to provide flexibility and adaptability under the code. This should be supported by a strong performance reporting regime, requiring signatories to regularly publish performance indicators and report on their progress.

Position 5

The ACMA expects the code will commit signatories to facilitate research, share relevant data, and undertake associated activities to improve understanding of misinformation in Australia. Platforms should consider ongoing avenues of collaboration between signatories, government, academia and other experts, and other relevant industries.

Position 6

The ACMA expects platforms to undertake an open, public consultation process when developing the code, with the code to be in place by no later than December 2020.

Position 7

The ACMA expects the code will require a robust, effective and accessible complaints handling regime. Users of digital platforms should also have access, free of charge, to an alternate dispute resolution process.

Position 8

The ACMA expects a representative body will be established to oversee the administration of the code. The code should also include a mandatory code review mechanism.