Facebook response to the Australian disinformation and misinformation industry code

MAY 2021

FACEBOOK

Summary

Facebook is pleased to be a signatory of the Australian disinformation and misinformation industry code, developed by the Digital Industry Group Inc (DIGI).

We have chosen to opt into every commitment under the code, and this response outlines in detail Facebook's commitments in the first year of the code. We have opted into the industry code for Facebook and Instagram and, although not required by the industry code, this response also includes information about steps that WhatsApp and Messenger have taken to address misinformation and disinformation.

As required by the code, we anticipate providing this information via an annual report that gives transparency to Australian policymakers and the community about the steps we are taking to combat disinformation and misinformation. Over time, we will endeavour to continue building out the information that we make available to support analysis and scrutiny about how best to combat misinformation and disinformation. In this way, we are able to track and monitor Facebook's progress over time.

We take a global approach to combatting misinformation and disinformation[1], and we constantly update our efforts in response to feedback, research, and changes in the nature of disinformation and misinformation. Our list of commitments span our global efforts, but also contain a number of new, Australia-specific commitments that go beyond what we have done in the past. In total, we are committing to **43 specific commitments** to meet the obligations specified in the voluntary industry code.

Some of the new, Australia-specific commitments include:

- We have prepared selected Australia-specific statistics about content on our platforms to encourage a sophisticated public policy debate about misinformation in Australia.
    - Since the beginning of the pandemic in March 2020 to the end of December 2020, globally we removed **over 14 million pieces of content** that constituted misinformation related to COVID-19 that may lead to harm, such as content relating to fake preventative measures or exaggerated cures. **110,000 pieces of content** were from Pages or accounts specific to Australia (noting that Australians benefitted from the content we removed from other countries as well).

[1] There is a significant amount of debate about the terms disinformation and misinformation. When we use these terms, throughout the submission, we use disinformation to refer to *behaviour* that is inauthentic (ie. people misrepresenting themselves in ways that are false) and misinformation refers to *content* that is false.

    - We have made a COVID-19 Information Centre available around the world to promote authoritative information to Facebook users. **Over 2 billion people globally** have visited the Information Centre; **over 6.2 million distinct Australians** have visited the Information Centre at some point over the course of the pandemic.

- We have offered additional support to the Australian Government to support authoritative information about the vaccine rollout, including a substantial provision of ad credits and the offer to build a vaccine finder service in Australia (as one of the first countries to receive this product outside the United States).
- We will be working towards expanding our fact-checking partner capability within Australia in 2021.
- We are undertaking public awareness campaigns in Australia to support Australians to spot and identify vaccine and COVID-19 misinformation online.
- Facebook will continue to fund Australia-specific research by independent experts and academics on media literacy, misinformation and disinformation in 2021. This builds on some of the research that we have already funded in the last twelve months, including:
    - providing funding (via the US-based National Association for Media Literacy Education) to support work done by academics from the Western Sydney University, Queensland University of Technology, and the University of Canberra to undertake Australia's first-ever nationwide adult media literacy survey. This was launched during a symposium in April 2021, held simultaneously across Sydney, Canberra and Brisbane to discuss misinformation and media literacy.
    - investing over US$2 million in a global round of funding for academic research on misinformation and polarisation. We announced the winners in August 2020, two of whom came from Australian universities.[2]
    - commissioning independent research by respected Australian academic Dr Andrea Carson to map government approaches to combating misinformation around the world, focussing on the Asia-Pacific region. The resulting report *Tackling Fake News* was launched in January 2021.
    - supporting a First Draft event held in March 2021 to discuss health misinformation

[2] A Leavitt, K Grant, 'Announcing the winners of Facebook's request for proposals on misinformation and polarization', https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/, 7 August 2020.

    - funding Dr Jake Wallis and the Australian Strategic Policy Institute to undertake a review of disinformation-for-hire, specifically targeting Australia and the Asia-Pacific region.

- Under our Australia-specific partnership with misinformation experts First Draft, we will be providing training material to influencers about preventing the amplification of misinformation, supporting an analytical paper on disinformation and misinformation amongst diaspora groups with a focus on Chinese language, and holding roundtables with experts.
- We are working with Australian Associated Press to develop a media literacy initiative for Australians about the importance of fact-checking and how to recognise and avoid the spread of misinformation.
- Facebook will undertake an initiative to support the provision of authoritative climate science information in Australia before the next report.
- Facebook will be extending the current industry-leading transparency we provide to political advertising to also cover social issues advertising.

The new Australia-specific commitments contained in this report are in addition to the significant global efforts that Facebook already undertakes to combat disinformation and misinformation. Our work includes:

- removing disinformation (in line with our Inauthentic Behaviour policy) and misinformation that can cause imminent, real-world harm
- paying independent, expert third-party fact-checkers to assess content on our services - and, if they find it to be false, we make significant interventions to limit the spread of that misinformation
- promoting authoritative information and labelling content to provide greater transparency to users
- encouraging industry-leading levels of transparency around political advertising
- building a publicly-available 'live dashboard' that allows anybody to publicly track and monitor COVID-19 public content on our services (in addition to making CrowdTangle freely available to journalists, third-party fact-checkers, and some academics).

Independent, expert academic research has found that the steps that Facebook is taking make a positive difference. Multiple studies have suggested that our efforts to

reduce misinformation on our platforms in recent years have had a meaningful impact.[3]

Facebook has supported new regulation for online content around the world, and we have supported the development of a voluntary industry code on disinformation since it was first proposed in Australia. We believe the first version of the code is a credible, world-leading first step in encouraging collaboration between the technology industry and governments to combat misinformation and disinformation. It learns the lessons of other efforts overseas (in particular, the EU Disinformation Code) to incorporate the best practices and address feedback from stakeholders to set a new benchmark of accountability for digital platforms such as Facebook in Australia.

We look forward to continuing to work with the Australian Communications and Media Authority, other government stakeholders, academics and experts to combat misinformation and disinformation in Australia.

[3] See the studies cited in T Lyons, 'New research showing Facebook making strides against false news', *Facebook Newsroom*, https://about.fb.com/news/2018/10/inside-feed-michigan-lemonde/.

List of Facebook commitments under the Australian voluntary industry code on disinformation and misinformation (May 2021)

Outcome 1: **Combatting misinformation / disinformation**	1. Facebook removes networks of accounts, Pages and Groups that violate our inauthentic behaviour policy, including disinformation that violates our policy on Coordinated Inauthentic Behaviour. 2. Facebook partners with experts and organisations who assist in providing tips or further investigation about possible inauthentic behaviour on our services. 3. Facebook removes misinformation that violates our Misinformation & Harm policy. 4. Facebook removes manipulated media, also known as “deepfakes”, that violates our Manipulated Media policy. 5. Facebook removes material that violates our Violence-Inducing Conspiracy Theory policy. 6. Facebook removes election-related misinformation that may constitute voter suppression. 7. Facebook removes fake accounts. 8. Facebook allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board. 9. Facebook partners with third-party fact-checking organisations, globally and in Australia, to assess the accuracy of content on our services. 10. Facebook will add additional fact-checking capability in Australia in 2021. 11. Facebook applies a warning label to content found to be false by third-party fact-checking organisations. 12. Facebook reduces the distribution of content found to be false by third-party fact-checking organisations. 13. Facebook proactively searches for content that makes claims debunked by our fact-checking partners, to apply the same treatments. 14. Facebook limits the ability to forward material via private messaging. 15. Facebook takes action on Pages, Groups, accounts or websites found to repeatedly share misinformation.

	16. Facebook removes Groups from recommendations if they violate our recommendation guidelines, including around misinformation. 17. Facebook makes available a detailed list of claims that we consider to violate our COVID-19 Misinformation & Harm policy. 18. Facebook makes information available via a dedicated website that outlines our efforts to combat misinformation. 19. Facebook makes on-platform reporting channels available to users for false information. 20. Facebook makes global transparent reports available regularly. 21. Facebook will supplement these reports with additional Australia-specific statistics, provided as part of this Annual Report process. 22. Facebook makes the service CrowdTangle freely available to journalists, third-party fact-checking partners, and some academics.
Outcome 2: **Disrupt monetisation and advertising incentives**	23. Facebook sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation.
Outcome 3: **Combat inauthentic user behaviour**	See items listed under Outcome 1.
Outcome 4: **Empower consumers to be informed**	24. Facebook provides contextual information around posts that users see from public Pages. 25. Facebook provides a COVID-19 Information Centre with verified, authoritative information about COVID-19. 26. Facebook will undertake an initiative to support the provision of authoritative climate science information in Australia before the next report. 27. Facebook uses in-product prompts to direct Australians to authoritative information on key topics.

	28. Facebook gives substantial ad credits to authoritative organisations, including the Australian Government and state and territory governments, to promote authoritative information. 29. Facebook directs users to authoritative information when they search for high-priority topics on Facebook. 30. Facebook directs users to authoritative information once they have seen or shared COVID-19 related misinformation. 31. Facebook will look for opportunities to continue to work with the Government on other ways to promote authoritative information. 32. Facebook promotes public service announcements to our users to encourage them to be wary of potential misinformation.
Outcome 5: **Political advertising**	33. Facebook requires all advertisers of political ads to complete an ad authorisation, which includes verifying the advertiser's identity. 34. Facebook requires political ads to include a disclaimer disclosing who is paying for the ad. 35. Facebook provides the Ad Library, a searchable archive of all political ads on our services in Australia, and will continue to add functionality to encourage scrutiny of political advertising. 36. Facebook enables an Ad Library report that provides aggregated spend information about Pages undertaking political ads. 37. Facebook will extend the policies and enforcement for political ads to social issue ads in 2021.
Outcome 6: **Research**	38. Facebook will continue to support research and events in relation to misinformation and media literacy. 39. Facebook will continue to support research and events in relation to disinformation. 40. Facebook provides a free CrowdTangle live display on COVID-19 publicly available to allow anybody to track public content on our platforms. 41. Facebook collaborates with researchers to undertake surveys of our users to assess their views on topics such as vaccines and climate change.

	42. Facebook provides data to researchers in a privacy-protective way via the Facebook Open Research and Transparency initiative.
Outcome 7: **Annual reports**	43. Facebook will continue to publish annual reports in Australia, such as these, to be transparent about the steps we are taking to combat disinformation and misinformation.

Table of contents

Facebook commitments

Outcome 1a
Signatories contribute to reducing the risk of Harms that may arise from the propagation of Disinformation and Misinformation on digital platforms by adopting a range of scalable measures.

Signatories will develop and implement measures which aim to reduce the propagation of and potential exposure of users of their services and products to Disinformation and Misinformation

Before providing more information about the steps we take to combat disinformation and misinformation, it is essential to define the terms. As Dr Andrea Carson says: "The lack of universally agreed definitions of terms such as online misinformation, disinformation and fake news presents significant obstacles to achieving consensus on how to tackle the problem. Even among experts [...], significant diversity of opinion emerged over the meanings of misinformation and disinformation."[4]

Throughout this report, when Facebook uses the terms "disinformation" and "misinformation", we mean:

- disinformation refers to *behaviour* that is inauthentic, with the intention to deceive, and
- misinformation refers to *content* that is false or misleading.

Facebook adopts a wide-ranging number of tactics to combat disinformation and reduce the spread of misinformation on our services. Broadly, these fall under a three part framework:

1. **remove** (1) disinformation; (2) misinformation that could cause imminent, real-world harm; (3) fake accounts, which can be vehicles for both disinformation and misinformation; and (4) allow for appeals in instances where we may not get this right
2. **reduce** the spread of other misinformation; and
3. promote authoritative information and develop tools to **inform** our users.

Specific commitments are listed under each of those headings.

[4] A Carson, *Fighting Fake News: A Study of Online Misinformation Regulation in the Asia-Pacific*, https://www.latrobe.edu.au/__data/assets/pdf_file/0019/1203553/carson-fake-news.pdf

Remove

We remove content that violates Facebook's policies set out in our Community Standards. Our Community Standards set the rules for what is and is not allowed on Facebook. More detail about the aspects of our Community Standards that relate to disinformation and misinformation are provided below. We also remove fake accounts which can be vehicles for disinformation or misinformation.

Disinformation

- **Facebook removes networks of accounts, Pages and Groups that violate our inauthentic behaviour policy, including disinformation that violates our policy on Coordinated Inauthentic Behaviour.**

  We note DIGI's recognition that there is no consensus on behaviour described as "disinformation". The closest term that Facebook uses in our own policies is inauthentic behaviour. In the social media landscape and beyond, foreign interference relies on inauthenticity - where users misrepresent themselves, through fake profiles or non-transparent behaviours - and coordination.

  We consider authentic communications as a central part of people's experience on Facebook. People find value in connecting with their friends and family, and they also find value in receiving updates from the Pages and organisations that they choose to follow. For this reason, authenticity has long been a requirement of Facebook's Community Standards.

  Our policies in this space have been through a number of iterations over recent years, to reflect our deepening understanding of the phenomenon of inauthentic behaviour.[5] We have an Inauthentic Behaviour policy, which has a number of components:

  - *Coordinated Inauthentic Behaviour* (CIB). We define this as groups of accounts and Pages that work together to mislead people about who they are and what they're doing. When we find domestic, non-government campaigns in which the use of fake accounts is central to the operation, we will remove all inauthentic and authentic accounts, Pages and groups directly involved in this activity.

  - *Foreign or Government Interference.* These are either (1) foreign-led efforts to manipulate public debate in another country in a way that is

[5] N Gleicher, 'How we respond to Inauthentic Behaviour - policy update', *Facebook Newsroom*, 21 October 2019, https://about.fb.com/news/2019/10/inauthentic-behavior-policy-update/

inauthentic; or (2) inauthentic behaviour operations run by a government to target its own citizens. If we see any of these instances, we will apply the broadest enforcement measures, including the removal of every on-platform property connected to the operation itself and the people and organisations behind it.

- *Other inauthentic behaviour*, including financially-motivated activity like spam or fake engagement tactics that rely on inauthentic amplification or evading use of enforcement (separate to use of fake accounts). The full list of tactics that we do not allow is available as part of our Community Standards.[6] We enforce against other inauthentic behaviour based on specific protocols that may involve temporary restrictions, warnings, down-ranking in Facebook News Feed, or removal.

We have invested significantly in a team that is able to detect the various forms of Inauthentic Behaviour on our services. In the first three months of 2021, we detected and disrupted 21 networks of CIB globally. We have previously detected and disrupted CIB networks targeting Australia, including two networks during the last 2019 election.[7]

- **Facebook partners with experts and organisations who assist in providing tips or further investigation about possible inauthentic behaviour on our services.**

  We are in an adversarial relationship with the bad actors who engage in Inauthentic Behaviour. We need to continually monitor their activity and adapt our enforcement, as they in turn change their tactics constantly to evade our detection. Our efforts to detect CIB are complemented by close relationships with experts and organisations in this space who can, on the basis of their expertise, provide us with tips or undertake further investigation of a particular phenomenon.

  For this reason, we maintain close relationships with a number of experts and organisations around the world who can assist. That includes Australian Government agencies, and the Australian Strategic Policy Institute (of which we are a major sponsor).

---

[6] Facebook, *Community Standards - Inauthentic Behaviour*, https://www.facebook.com/communitystandards/inauthentic_behavior

[7] Facebook, *Facebook's submission to the Joint Standing Committee on Electoral Matters*, 8 October 2019, https://www.aph.gov.au/DocumentStore.ashx?id=77c96820-3a40-4670-afba-3677e861a93d&subId=671217.

Misinformation

- **Facebook removes misinformation that violates our Misinformation & Harm policy.** We have had a policy on Misinformation and Harm since 2018, which has been used in instances such as when we removed harmful health misinformation during the measles outbreak in Samoa towards the end of 2019. When COVID-19 was declared a global health emergency in January 2020, we worked with experts around the global - in particular, the World Health Organization - to identify COVID-related claims that could cause imminent, physical harm.

  Originally, this included COVID-19 claims such as false claims about:
  - the existence or severity of COVID-19
  - transmission and immunity - including claims that COVID-19 is transmitted via 5G
  - guaranteed cures or prevention methods
  - discouraging good health practices
  - access to essential health services.

  Since the beginning of the pandemic to the end of 2020, we removed over 14 million claims globally that fell under these categories.

  In December 2020, we expanded our policy to cover false claims about COVID vaccines.[8]

  In February, we expanded it to include claims about vaccines generally. That means we are now removing false claims about:
  - vaccines causing autism
  - vaccines causing Sudden Infant Death Syndrome
  - vaccines causing or encouraging the disease they are intended to prevent
  - natural immunity being more effective than vaccines
  - the timing of vaccines (ie. that receiving vaccines in short succession is dangerous)
  - vaccines not being effective against the disease they are intended to prevent
  - acquiring measles not causing death
  - natural immunity being safer than vaccine-acquired immunity
  - Vitamin C is as effective as vaccines in preventing diseases for which vaccines exist.

[8] Facebook Newsroom, Removing false claims about COVID-19 Vaccines, December 2020 https://about.fb.com/news/2020/12/coronavirus/#removing-covid-vaccine-misinformation

We remove these claims when we become aware of them.

Each policy change is underpinned by extensive consultation with experts, to ensure we are taking a proportionate approach to considering the harm that can arise from a piece of misinformation. We continue to review these policies and update them regularly (for example, in the last two months, we included an amendment to the policy to indicate that false claims about blood clots as a side-effect of vaccines will be removed, if it relates to a vaccine with no scientific concerns about blood clots).

- **Facebook removes manipulated media, also known as "deepfakes", that violates our Manipulated Media policy.**

  After consulting with more than 50 global experts with technical, policy, media, legal, civic and academic backgrounds, we announced in 2020 that we will be removing manipulated media if: (1) it has been edited or synthesised – beyond adjustments for clarity or quality – in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say; and (2) it is the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic.[9]

- **Facebook removes material that violates our Violence-Inducing Conspiracy Theory policy.**

  In August and October 2020, we expanded our dangerous organisations policy to capture content relating to "violence-inducing conspiracy theories". Prior to this change, we had already been removing conspiracy theory material that advocated violence. Following these changes, we now remove Facebook Pages, Facebook Groups and Instagram accounts associated with QAnon.[10] As of January 2021, we had removed about 3,300 Pages, 10,500 groups, 510 events, 18,300 Facebook profiles and 27,300 Instagram accounts for violating our policy against QAnon. Some of these Pages, groups, events, profiles and accounts were located in Australia.

[9] M Bickert, *Enforcing Against Manipulated Media*, https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/, 6 January 2020.

[10] Facebook, 'An update to how we address movements and organizations tied to violence', *Facebook Newsroom*, blog post updated 19 January 2021, https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.

We are continuing with consultation and consideration of a potential harmful conspiracy theory policy that accounts for harms broader than violence, as advocated by QAnon.

- **Facebook removes election-related misinformation that may constitute voter fraud and/or interference.** Under our policies, we prohibit: misrepresentation of the dates, locations, times, and methods of voting or voter registration (for example: claims that you can vote using an online app); misrepresentations of who can vote, how to vote, qualifications for voting and whether a vote will be counted; or misrepresentation of who can vote, qualifications for voting, whether a vote will be counted, and what information or materials must be provided in order to vote;. We also do not allow statements that advocate, provide instructions, or show explicit intent to illegally participate in a voting process.

  Voting is essential to democracy, which is why we take a stricter policy on misrepresentations and misinformation that could result in voter fraud or interference.

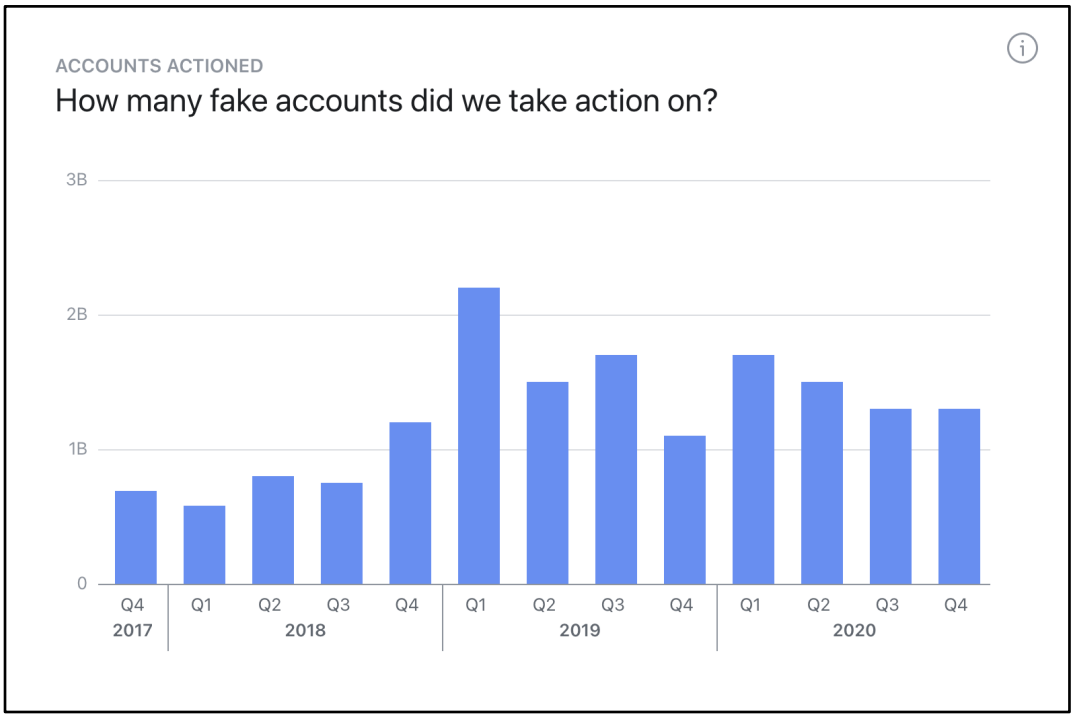
Fake accounts

- **Facebook removes fake accounts.** Fake accounts can often be the vehicle for harmful content, including misinformation. We block billions of fake accounts at creation, and also removed 1.3 billion fake accounts between October and December 2020, the majority of these accounts were caught within minutes of registration.[11] Of these, 99.6 per cent of these accounts were detected proactively via artificial intelligence, before they were reported to us. There has been a general decline since Q1 2019 in the volume of fake accounts we have been detecting, as our ability to detect and block attempts to create fake accounts at creation have been improving.

  We report on these figures quarterly through our Community Standards Enforcement Report, a voluntary transparency effort that allows for scrutiny of our efforts to enforce Facebook and Instagram's community standards.[12]

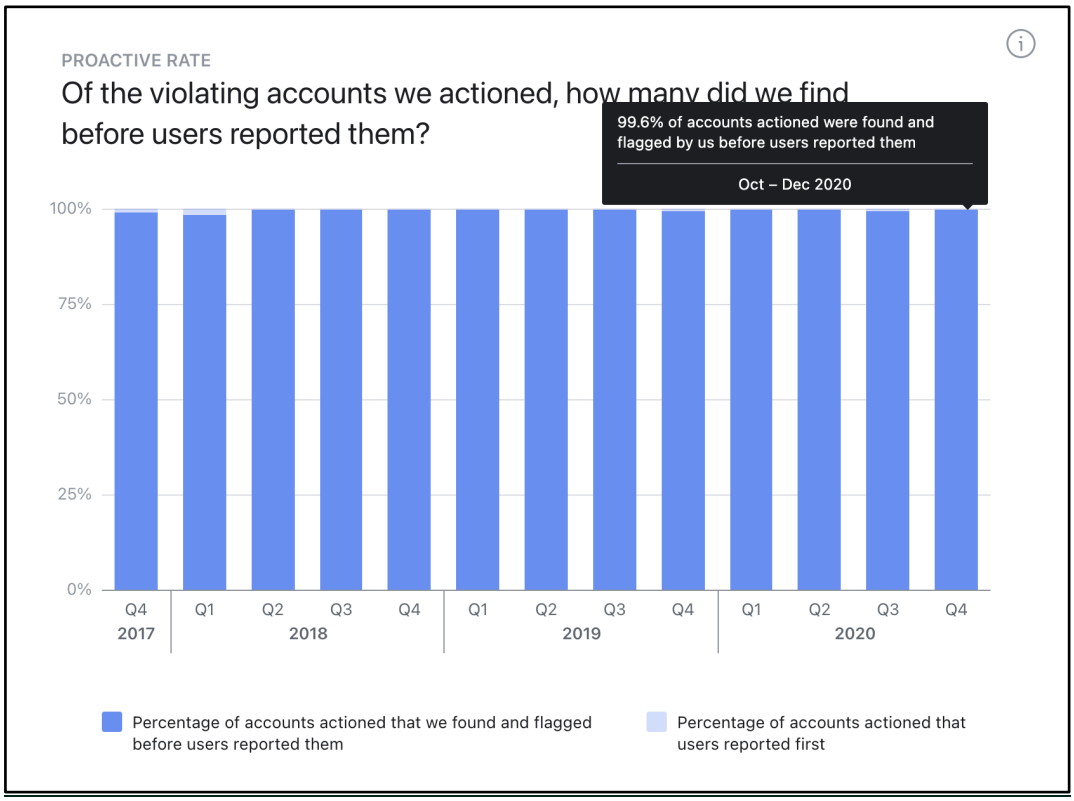
[11] Facebook, *Community Standards Enforcement Report*, https://transparency.facebook.com/community-standards-enforcement#fake-accounts
[12] ibid.

Figure 1: How many fake accounts did we take action on?



Source: Community Standards Enforcement Report, Q4 2020

Figure 2: Of the violating accounts we actioned, how many did we find before users reported them?



Source: Community Standards Enforcement Report, Q4 2020

Appeals

- **Facebook allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.** We recognise that our Community Standards may not always get it right: we may enforce our policies incorrectly, or reasonable minds may differ in terms of where we should draw the line. For that reason, we allow for appeals of decisions we make in relation to our Community Standards.

  In order to help give policymakers and the community confidence about our content governance, we have established the Oversight Board. The members of the Oversight Board were appointed in May 2020[13] and they began taking cases in October 2020.[14]

  The Oversight Board was borne out of the recognition that critical decisions about content should not be left to companies alone. Content decisions can have significant consequences for free expression and companies like Facebook - notwithstanding our significant investments in detection, enforcement and careful policy development - will not always get it right.

  The Oversight Board comprises 20 experts in human rights and technology - including the Australian academic Professor Nic Suzor - and will increase over time to 40 members. The Board is entirely independent and hears appeals on Facebook's decisions relating to content on Facebook and Instagram (beginning with decisions where content was removed). We have agreed that the Board's decisions will be binding, and the Board is also able to make recommendations about Facebook's policies.[15]

  The Oversight Board has begun issuing its decisions from January 2021.[16] This included a decision and policy recommendation related to our misinformation and harm policies, specifically how we treat claims about hydroxychloroquine. And, from 13 April 2021, users are now able to appeal content that we have *not*

[13] N Clegg, 'Welcoming the Oversight Board', *Facebook Newsroom*, 6 May 2020, https://about.fb.com/news/2020/05/welcoming-the-oversight-board/
[14] B Harris, 'Oversight Board Selects First Cases to Review', *Facebook Newsroom*, 1 December 2020, https://about.fb.com/news/2020/12/oversight-board-selects-first-cases-to-review/
[15] B Harris, 'Establishing structure and governance for an independent oversight board', *Facebook Newsroom*, 17 September 2019, https://about.fb.com/news/2019/09/oversight-board-structure/
[16] M Bickert, 'Responding to the Oversight Board's First Decisions', *Facebook Newsroom*, 28 January 2021, https://about.fb.com/news/2021/01/responding-to-the-oversight-boards-first-decisions/

taken down (ie. content that we have left up that they believe should be removed).[17]

We believe the Oversight Board is a significant innovation in content governance and a first-of-its-kind initiative. It will make Facebook more accountable for our content decisions and will help to improve our decision-making.

Reduce

For content that does not violate our Community Standards but is rated as false by Facebook's independent third-party fact-checking partners, we significantly reduce the number of people who see it. We believe that public debate and democracy are best served by allowing people to debate different ideas, even if they are controversial or wrong - but we take steps to limit the distribution of misinformation that has been found to be false by independent, expert fact checkers.

- **Facebook partners with third-party fact-checking organisations, globally and in Australia, to assess the accuracy of content on our services.** We have commercial arrangements with independent third-party fact-checking organisations for them to review and rate the accuracy of posts on Facebook and Instagram. In Australia, we partner with Australian Associated Press and Agence France Presse, both certified by the non-partisan International Fact-Checking Network, as part of a network of over 80 fact-checking partners around the world covering more than 60 languages.

  All fact-checks by these partners are publicly available on their websites.[18]

  In addition to our existing commercial arrangements, we have given two $1 million grants to fact-checking partners to improve their capacity during the high-volume time associated with COVID-19, including in relation to misinformation on WhatsApp.[19]

- **Facebook will add additional fact-checker capability in Australia in 2021.**

---

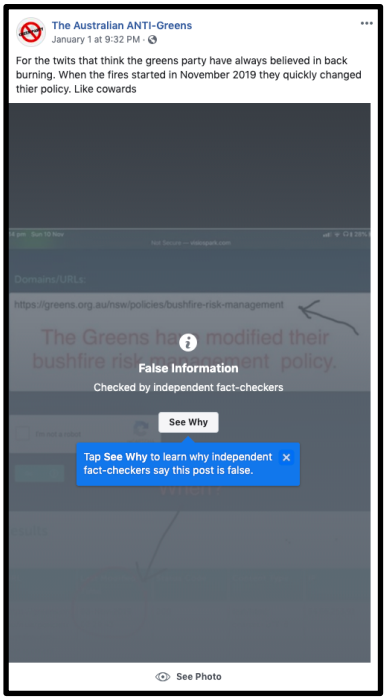
[17] G Rosen, 'Users Can Now Appeal Content Left Up on Facebook or Instagram to the Oversight Board', https://about.fb.com/news/2021/04/users-can-now-appeal-content-left-up-on-facebook-or-instagram-to-the-oversight-board/, 13 April 2021.

[18] Agence France Presse Australia, *Fact Check*, https://factcheck.afp.com/afp-australia; Australian Associated Presse, *AAP Fact Check*, https://www.aap.com.au/category/factcheck/

[19] Supporting Fact-checkers and Local News Organisations https://about.fb.com/news/2020/12/coronavirus/#supporting-fact-checkers

Facebook is working towards expanding our fact-checking capacity in Australia in 2021, including by establishing an additional independent partner. This commitment is subject to willingness and capacity by an International Fact-Checking Network (IFCN)-certified fact-checker to participate in our third-party fact-checking program.

- **Facebook applies a warning label to content found to be false by third-party fact-checking organisations.** Once a third-party fact-checking partner rates a post as 'false', we apply a warning label that indicates it is false and shows a debunking article from the fact checker. It is not possible to see the content without clicking past the warning label. When people see these warning labels, 95% of the time they do not go on to view the original content.



We also apply the label in instances where people go to share a claim that has been found to be false. When someone tries to share a post that's been rated by a fact-checker, we'll show them a pop-up notice so people can decide for themselves what to read, trust, and share.

- **Facebook reduces the distribution of content found to be false by third-party fact-checking organisations.** We reduce the distribution of the content so it appears lower in News Feed, which slows its distribution significantly. And on Instagram, we remove it from Explore and hashtag pages and downrank content in Feed and Stories.

- **Facebook proactively searches for content that makes claims debunked by our fact-checking partners, to apply the same treatments.** Based on one fact-check, we're able to kick off similarity detection methods that identify duplicates of debunked stories. Using this technology, we are able to limit the distribution of similar posts: in April 2020 alone, we applied the label and reduced the distribution of more than 50 million posts worldwide, based on more than 7,500 fact-checks. And, between March and October 2020, we put these warning labels on 167 million pieces of content.

- **Facebook limits the ability to forward material via private messaging.** We have instituted strict forward limits for messages to help users identify when a message was not written by the sender and to introduce friction in the experience to constrain message virality.

  On WhatsApp, you can only forward a message to five additional chats at one time, making it one of the few services to intentionally limit sharing.[20] This change reduced the amount of forwarded messages on WhatsApp by more than 25%.

  In 2020, the service set stricter limits for messages that have been forwarded many times. These messages are marked with double arrows and labelled as 'Forwarded many times'. WhatsApp has introduced a limit to the forwarding of these messages which means that they can only be forwarded to one other chat at a time. Although highly forwarded messages make up a very small percentage of all messages sent on WhatsApp, the introduction of this limit has further reduced the number of these messages by over 70% globally.

  On Messenger, we introduced a similar forwarding limit in September 2020.[21] Messages can now only be forwarded to five people or Groups at a time.

- **Facebook takes action on Pages, Groups, accounts, or websites found to repeatedly share misinformation.** When Pages or websites repeatedly share content that's been debunked by fact-checking partners, they will see their overall distribution reduced, and will lose the ability to advertise or monetise within a given time period. If they continue to share misinformation beyond that, their Page is removed in its entirety.

  This includes Pages operated by public figures.

---

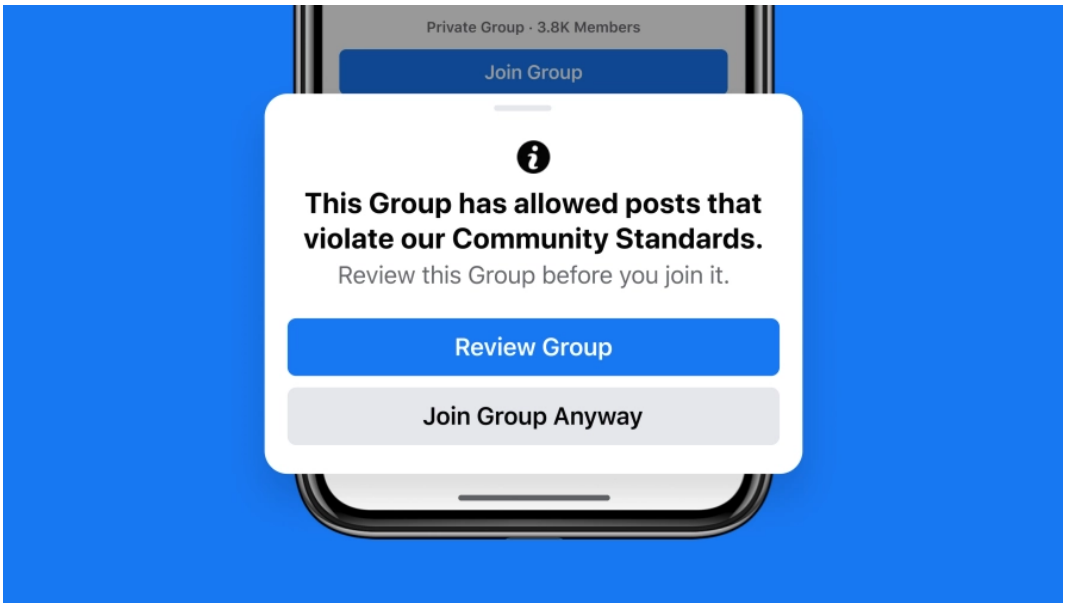
[20] WhatsApp, *About forwarding limits*, https://faq.whatsapp.com/general/chats/about-forwarding-limits

[21] J Sullivan, 'Introducing a forwarding limit on Messenger', *Facebook Newsroom*, https://about.fb.com/news/2020/09/introducing-a-forwarding-limit-on-messenger/, 3 September 2020.

- **Facebook removes Groups from recommendations if they violate our recommendation guidelines, including around misinformation.** People turn to Facebook Groups to connect with others who share their interests, but even if they decide to make a group private, they have to play by the same rules as everyone else. Our Community Standards apply to public and private groups, and our proactive detection tools work across both.

  We are continually taking steps to keep Groups safe and reduce the spread of harmful content like misinformation. Some of the steps we have announced in the last 12 months include:
  - preventing admins and moderators of groups taken down for policy violations from creating *any* new Groups for a period of time[22]
  - archiving Groups that have been without an admin for a period of time[23]
  - removing health groups from recommendations (discussed in more detail below)
  - informing people when a Group they administer is sharing misinformation: Group admins are also notified each time a piece of content rated false by fact-checkers is posted in their group, and they can see an overview of this in the Group Quality tool
  - removing civic and political Groups from recommendations. This has occurred originally in the US, but we are now beginning to also roll it out globally[24]
  - we have just started to let people know when they're about to join a group that has Community Standards violations, so they can make a more informed decision before joining.[25]

[22] T Alison, 'Our latest steps to keep Facebook Groups safe', *Facebook Newsroom*, 17 September 2020, https://about.fb.com/news/2020/09/keeping-facebook-groups-safe/.
[23] ibid.
[24] T Alison, 'Changes to keep Facebook Groups safe', *Facebook Newsroom*, 17 March 2021.
[25] ibid.



- limiting invite notifications for Groups with Community Standards violations, so people are less likely to join.[26]
- for existing members of a Group with Community Standards violations, we are reducing the distribution of that Group's content so that it's shown lower in News Feed.[27]
- when someone has repeated violations in Groups, we will block them from being able to post or comment for a period of time in any group. They also won't be able to invite others to any groups, and won't be able to create new groups.[28]

One of the most important steps we take to keep Groups safe relates to our recommendations. Across our apps, we make recommendations to help users discover new communities and content they are likely to be interested in. We suggest Pages, Groups, Events and more based on content that users have expressed interest in and actions you take on our apps - and these recommendations are personalised.

Since recommended content doesn't come from accounts you choose to follow, it's important that we have certain standards for what we recommend. This helps ensure we don't recommend potentially sensitive content to those who don't explicitly indicate that they wish to see it. It helps to ensure we are reducing the distribution of content that people tell us they don't want to see.

[26] ibid.
[27] ibid.
[28] ibid.

In August 2020, we released our first set of recommendation guidelines.[29] These detail the types of content that we allow on our platform, but we do not promote via recommendations. This includes:

- Sensitive or low-quality content about health or finance, such as "miracle cures".
- Designated types of false or misleading content, such as claims that have been debunked by third-party fact-checkers.

[29] G Rosen, 'Recommendation Guidelines', *Facebook Newsroom*, 31 August 2020, https://about.fb.com/news/2020/08/recommendation-guidelines/.

Inform

We take a number of steps in order to inform users about the content that they see on Facebook. These are outlined in more detail under Outcome 4. At a high level, they include:

- providing contextual information around posts that users see from public Pages, such as the context button, the breaking news tag, notifications for news articles that are more than 90 days old, labels for posts from state-controlled media, and notifications to give context around COVID-19 related links. This is in addition to the transparency we provide around Facebook Pages more generally.

- providing a COVID-19 Information Centre with verified, authoritative information about COVID-19.

- launching an initiative to support the provision of authoritative climate science information in Australia before the next report

- using in-product prompts to direct Australians to authoritative information on key COVID-19 related topics.

- providing significant amounts of ad credits to authoritative organisations, including the Australian Government and state and territory governments, to promote authoritative information.

- directing users to authoritative information when they search for high-priority topics on Facebook, such as ‘coronavirus’ and harmful misinformation (like the QAnon conspiracy theory).

- directing users to authoritative information once they have seen or shared COVID-19 related misinformation.

- looking for opportunities to continue to work with the Australian Government on other ways to promote authoritative information.

- promoting public service announcements to our users to encourage them to be wary of potential misinformation.

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

Signatories will implement and publish policies and procedures and any appropriate guidelines or information relating to the prohibition and/or management of user behaviours that may propagate Disinformation and Misinformation via their services or products.

We recognise that users need to be provided with information about the types of behaviour and content that is not permitted on Facebook. We are also conscious not to provide too much detail which would allow bad actors to skirt our policies and evade enforcement.

As outlined above under Outcome 1a, Facebook maintains a number of policies under our Community Standards to remove harmful misinformation and disinformation. These include policies related to:

- misinformation that could cause imminent, physical harm
- manipulated media
- violence-inducing conspiracy theories
- election-related misinformation that may constitute voter fraud and/or interference
- fake accounts
- coordinated inauthentic behaviour.

As with all of our Community Standards, these are available at facebook.com/communitystandards.

- **Facebook makes available a detailed list of claims that we consider to violate our COVID-19 Misinformation & Harm policy**.[30] We made this publicly available in February 2021, following feedback from the Oversight Board that Facebook should be more transparent about the specific claims that we consider to represent misinformation that could cause imminent, physical harm. In a dedicated part of our Community Standards, we provide a detailed list in order to be more transparent about our policies. These are available at: https://www.internmc.facebook.com/help/230764881494641

[30] Available at https://www.internmc.facebook.com/help/230764881494641

- **Facebook makes information available via a dedicated website that outlines our efforts to combat misinformation.** Facebook has provided a large number of updates about our policies and procedures relating to misinformation via our Newsroom (about.fb.com/news), especially over the last twelve months of the pandemic. Some of these updates include:
    - 8 February 2021: Reaching billions of people with COVID-19 vaccine information
    - 18 December 2020: updating our ads policies relating to COVID-19 vaccines
    - 3 December 2020: removing false claims about COVID-19 vaccines
    - 15 July 2020: launching a new part of the COVID-19 Information Centre called "Facts About COVID"
    - 16 April 2020: Taking additional steps to limit COVID-19 related misinformation
    - 14 April 2020: Helping the World Health Organization share authoritative information via Messenger
    - 7 April 2020: Helping people get reliable information in Groups and Events
    - 26 March 2020: Launching the Messenger COVID Coronavirus Hub
    - 25 March 2020: Combatting coronavirus misinformation across our apps
    - 18 March 2020: Launching the COVID-19 Information Centre on Facebook
    - 17 March 2020: Supporting fact-checkers and local news organisations
    - 13 March 2020: Connecting people with credible health information on Instagram
    - 6 March 2020: Removing COVID-19 Misinformation on Instagram
    - 26 February 2020: Connecting people to accurate and authoritative health resources
    - 30 January 2020: Limiting misinformation and harmful content.

  In order to help users understand these new policy updates collectively, and how they relate to misinformation efforts that were in place prior to COVID-19, we have been a dedicated website that summarises our work to combat misinformation in a clear and user-friendly way. This is available at: https://www.facebook.com/combating-misinfo

Outcome 1c

Users can report content and behaviours to Signatories that violates their policies under 5.10 through publicly available and accessible reporting tools.

Signatories will implement and publish policies, procedures and any appropriate guidelines or information regarding the reporting of the types of content and behaviours that may propagate Disinformation and Misinformation via their platforms.

We outline information about the policies we apply for misinformation and disinformation publicly, to be clear to our users in how we apply our policies.

We also provide avenues for users to tell us that they consider a piece of content to be false news. We use the term 'false news' because we have found it is easier for users to understand than terms such as "misinformation" or "disinformation". As Dr Andrea Carson says, even experts are not able to agree on use of terms such as misinformation and disinformation.[31]

Reports against false news provides a signal to us about content that users consider to be misinformation. Disinformation is more complex. Users are not able to discern if material constitutes disinformation because the definition of disinformation relies on the *behaviour* of the actors spreading the material, not the nature of the *content*.

Actors engaged in disinformation need not necessarily use misinformation; most of the content shared by coordinated manipulation campaigns isn't provably false, and would in fact be acceptable political discourse if it was shared by authentic actors. The real issue is that the actors behind these campaigns are using deceptive behaviours to conceal the identity of the organisation behind a campaign, make the organisation or its activity appear more popular or trustworthy than it is, or evade our enforcement efforts.

It is not possible for an average user to discern whether a piece of content is part of a disinformation campaign, given the falsity arises from the actor's behaviour, rather than the content.

That is why we rely on a range of other signals, including referrals from external experts, to help us identify disinformation. This provides a more reliable signal than providing reporting categories for "CIB" or "disinformation" on-platform. However,

[31] A Carson, *Fighting Fake News: A Study of Online Misinformation Regulation in the Asia-Pacific*, https://www.latrobe.edu.au/__data/assets/pdf_file/0019/1203553/carson-fake-news.pdf

reports that users submit for “fake accounts” or “false information” may still constitute disinformation for the purposes of this code.

- **Facebook makes on-platform reporting channels available to users for false information.** We make available reporting to users on-platform. On both Facebook and Instagram, users are able to report on-platform by clicking on the three dots at the top right of any post, and select “False information”, which is a dedicated reporting category for misinformation. There are sub-categories provided for health, politics, social issues, or other issues. On Facebook, reporting is also possible for Pages and Groups.

  We do not recommend relying on reporting channels that are not on-platform because (1) we may not be able to identify the content that’s the subject of the report; and (2) on-platform reporting is more simple and timely than asking users to leave their user experience in order to report something they see on our services.

Outcome 1d
Users will be able to access general information about Signatories' actions in response to reports made under 5.11.

Signatories will implement and publish policies, procedures and/or aggregated reports (including summaries of reports made under 5.11) regarding the detection and removal of content that violates platform policies, including but not necessarily limited to content on their platforms that qualifies as Misinformation and/or Disinformation.

We make aggregated reports publicly available about our misinformation and disinformation efforts to give users, policymakers, researchers, experts and the broader community information about the effectiveness of our detection and enforcement. Specifically:

- **Facebook makes global transparent reports available regularly.** Across all of our enforcement efforts, Facebook has worked hard to develop industry-leading reports that encourage transparency and accountability of our efforts to combat harmful content. These include:
    - our Community Standards Enforcement Report, which provides quarterly statistics about our efforts to combat a range of types of harmful content, for example, fake accounts. These are available at: https://transparency.facebook.com/community-standards-enforcement
    - reports on Coordinated Inauthentic Behaviour. Every month, we publish information about the CIB networks that we have detected and disrupted in the previous month. These are available at: https://about.fb.com/news/tag/coordinated-inauthentic-behavior/

- **Facebook will supplement these reports with additional Australia-specific statistics, provided as part of this Annual Report process.** In 2021, we will add one additional reporting timeframe in order to assist the ACMA with assessing progress under this code.

  To that end, we are pleased to release the following new statistics:
    - Since the beginning of the pandemic (March 2020) through to January 2021, globally we removed **over 14 million pieces of content** that constituted misinformation related to COVID-19 that may lead to harm, such as content relating to fake preventative measures or exaggerated cures. **110,000 pieces of content** were from Pages or accounts specific

to Australia (noting that Australians also benefitted from material we removed from other countries as well).
    - We have made a COVID-19 Information Centre available around the world to promote authoritative information to Facebook users. **Over 2 billion people globally** have visited the Information Centre; **over 6.2 million distinct Australians** have visited the Information Centre at some point over the course of the pandemic.

- **Facebook makes the service CrowdTangle freely available to journalists, third-party fact-checking partners and some academics.** CrowdTangle is an insights service that makes it easy to follow, analyse and report on what's happening with public content on social media. It covers content on Facebook, Instagram and Reddit and Facebook makes this tool available for free.
  Many journalists and researchers use this tool to scrutinise our efforts in relation to misinformation and disinformation. Further information is provided under Outcome 6 on a CrowdTangle Live Display that we have made publicly available to everybody relating to COVID-19.

Outcome 2
Advertising and/or monetisation incentives for Disinformation are reduced.

Signatories will implement policies and processes that aim to disrupt advertising and/or monetisation incentives for Disinformation.

Facebook already has a number of policies and processes in place to demonetise or disrupt advertising for bad actors.

In short, if any user is engaged in Coordinated Inauthentic Behaviour, they are suspended from our services entirely - including advertising. For other users engaged in spreading disinformation or misinformation, we take other action, described in more detail below.

- **Facebook sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation.** Facebook requires that any users who advertise on our services comply with our Advertising Policies. One of the key obligations in our Advertising Policies is that ads must comply with our Community Standards: this means that any content that violates our Community Standards will also violate our Advertising Policies. We will remove those ads.

  But our Advertising Policies go further than our Community Standards and prohibit ads that may otherwise be allowed in an organic, non-paid sense. We strongly believe that the standards for advertising should be stricter than for organic, non-paid content because: (1) it is possible to extend the reach and distribution of that content; and (2) users are only able to see organic content if they opt in (for example, connect with a friend or choose to follow a Page). Ads are one of the few ways to reach a user without them opting in, and so it is correct to set a higher standard.

  Some of the steps we take include:

  - We do not allow advertising of any content that has been found to be false by an independent third-party fact-checker, even though these posts are allowed to be shared on Facebook in an organic, non-paid way.

  - We also take a graduated enforcement approach to misinformation. If users repeatedly share misinformation, we may take steps such as removing their Page from recommendations. Users who persist in

violating our policies will have their ability to advertise or monetise wholesale removed. Users who continue to share misinformation after that are removed from Facebook or Instagram entirely.

- For advertisers who want to run ads about elections or politics targeting Australia, we require them to complete an ad authorisation process and labelling their ads with a publicly visible disclaimer indicating who has paid for the ad.

- We invest significant resources in ensuring the integrity of ads and detecting harmful or inauthentic behaviours in ads. For example, we have invested significantly in being able to combat the technique of "cloaking".[32]

[32] R Leathern, 'Addressing cloaking so people see more authentic posts', *Facebook Newsroom*, 9 August 2017, https://about.fb.com/news/2017/08/news-feed-fyi-addressing-cloaking-so-people-see-more-authentic-posts/.

Outcome 3

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

Signatories commit to take measures that prohibit or manage the types of user behaviours that are designed to undermine the security and integrity of their services and products, for example, the use of fake accounts or automated bots that are designed to propagate Disinformation.

As outlined under Outcome 1a, Facebook takes a number of actions against inauthentic user behaviours. This includes:

- Detection and removal of billions of fake accounts every quarter.
- Detecting and disrupting networks of Inauthentic Behaviour on our services.
- Partnering with experts and organisations who assist in providing tips or further investigation about possible coordinated inauthentic behaviour on our services.

Outcome 4

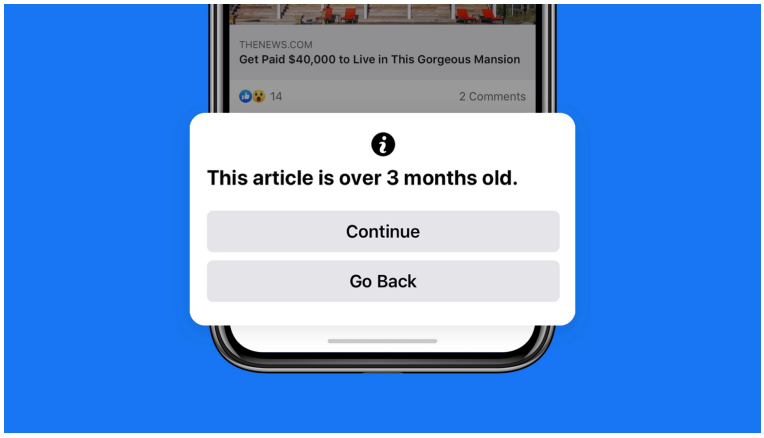
Users are enabled to make more informed choices about the source of news and factual content accessed via digital platforms and are better equipped to identify Misinformation.

Signatories will implement measures to enable users to make informed choices about news and factual information and to access alternative sources of information.

We take a number of steps in order to inform users about the content that they see on Facebook. In particular, we have gone further to actively promote authoritative information around COVID-19.

- **Facebook provides contextual information around posts that users see from public Pages.**

  We have developed a number of other labels and signals for users relating to the trustworthiness of posts they see on Facebook. These include:
  - the context button, which provides information about the sources of articles in News Feed[33]
  - the breaking news tag, to help people easily identify timely news or urgent stories[34]
  - a new notification screen that we recently introduced that lets people know when news articles they are about to share are more than 90 days old.[35]

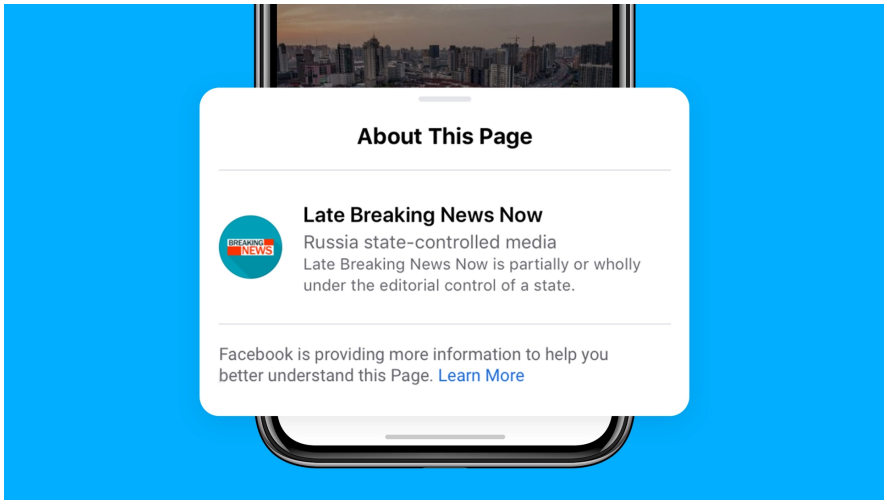


[33] J Smith, A Leavitt & G Jackson, 'Designing New Ways to Give Context to Stories', *Facebook Newsroom*, https://about.fb.com/news/2018/04/inside-feed-article-context/, 8 April 2018

[34] J Rhyu, 'Enabling Publishers to Label Breaking News on Facebook', *Facebook for Journalism*, 6 April 2020, https://www.facebook.com/journalismproject/facebook-breaking-news-label.

[35] J Hegeman, 'Providing people with additional context about content that they share', *Facebook Newsroom*, 25 June 2020, https://about.fb.com/news/2020/06/more-context-for-news-articles-and-other-content/.

We have developed labels for particular situations where experts have advised us there is particular risk. For example, since the middle of 2020, we have had a specific label for content from news organisations that are partially or fully under the control of their government.[36] This helps users to understand if a news publication that they read may be under the influence of a particular government. These are now available on both Facebook and Instagram.



We have started rolling out a new notification to give people more context about COVID-19 related links when they are about to share them. The notification will help people understand the recency and source of the content before they share it. It will also direct people to our COVID-19 Information Centre to ensure people have access to credible information about COVID-19 from global health authorities.[37]

These are all just-in-time notices that provide labelling and information related to specific pieces of content that individuals have seen.

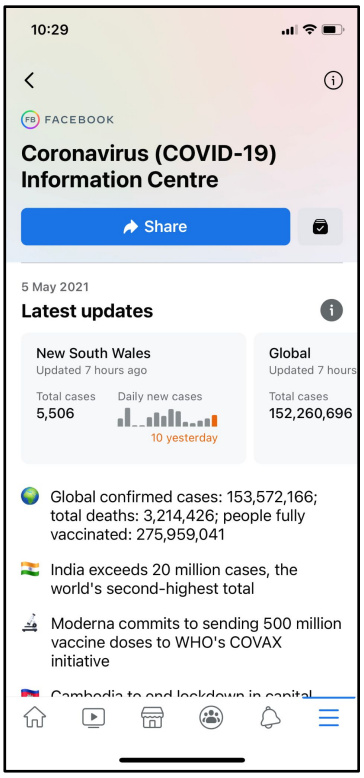
We also provide information about Pages more generally, for those users who would like more information around the author behind the content.[38] Every Page contains a Page Transparency tool, which includes information such as:

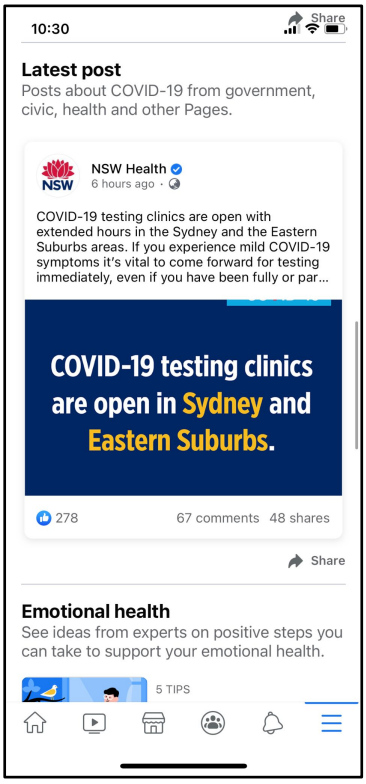
- previous names for the Page
- the number of admins for the Page

[36] N Gleicher, 'Labeling State-Controlled Media on Facebook', *Facebook Newsroom*, 4 June 2020, https://about.fb.com/news/2020/06/labeling-state-controlled-media/.
[37] J Hegeman, 'Providing people with additional context about content that they share', *Facebook Newsroom*, 25 June 2020, https://about.fb.com/news/2020/06/more-context-for-news-articles-and-other-content/.
[38] R Leathern and E Rogers, 'A New Level of Transparency for Ads and Pages', *Facebook Newsroom*, 28 June 2018, https://about.fb.com/news/2018/06/transparency-for-ads-and-pages/.

- in which country the Page admins are located, and
- ads currently being run by the Page.

- **Facebook provides a COVID-19 Information Centre with verified, authoritative information about COVID-19.** We have launched a Coronavirus Information Centre on Facebook in Australia that provides a centralised hub of the most up-to-date information on COVID-19, including diagnoses numbers etc official Australian Government information, access to authoritative health resources, and curated news sources. We send regular alerts to those who are subscribed to the Coronavirus Information Centre so they are aware of timely updates.





So far, more than 2 billion Facebook users around the world have accessed the COVID-19 Information Centre. This includes **over 6.2 million distinct Australians** have visited the Information Centre at some point over the course of the pandemic.

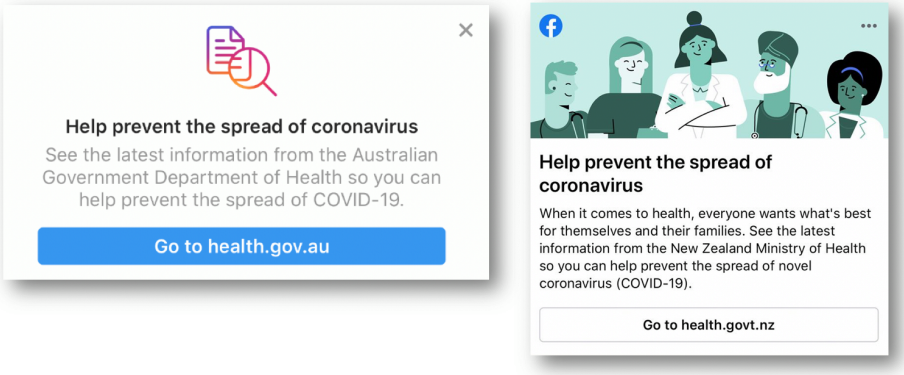
Throughout the pandemic, we have adapted the COVID-19 Information Centre to contain curated, factual information most relevant at that time. At the time of writing, there are modules about: vaccines testing (which have been developed with the World Health Organization); looking after your emotional and mental health during the pandemic; myth-busting about COVID-19; prevention tips; and combatting domestic and family violence. It also connected Australians to content from hand-selected trusted organisations,

like the Australian Government, state and territory governments, and UNICEF. This helps to extend the reach and impact of authoritative communications from those trusted sources.

- **Facebook will undertake an initiative to support the provision of authoritative climate science information in Australia before the next report**. We have already announced our intentions globally to support authoritative information about climate science and will be undertaking an Australia initiative before the next report.[39]

- **Facebook uses in-product prompts to direct Australians to authoritative information on key COVID-19 related topics.** Since the very beginning of the crisis, we have been displaying on Facebook and Instagram prompts to direct users to official sources of information, including from the Australian Government and the World Health Organization. These have been seen by every Facebook and Instagram user in Australia multiple times, either in their Feeds or when they search for coronavirus-related terms. In the last month, we ran prompts in Australia to urge people to wear a mask while outside at all times.

  Globally, more than 600 million people around the world have clicked on these prompts to learn more.



We have also used similar techniques to promote authoritative information from the Australian Electoral Commission in the last Federal election. Hundreds of thousands of Australians clicked through to authoritative information from the AEC via one of these prompts.

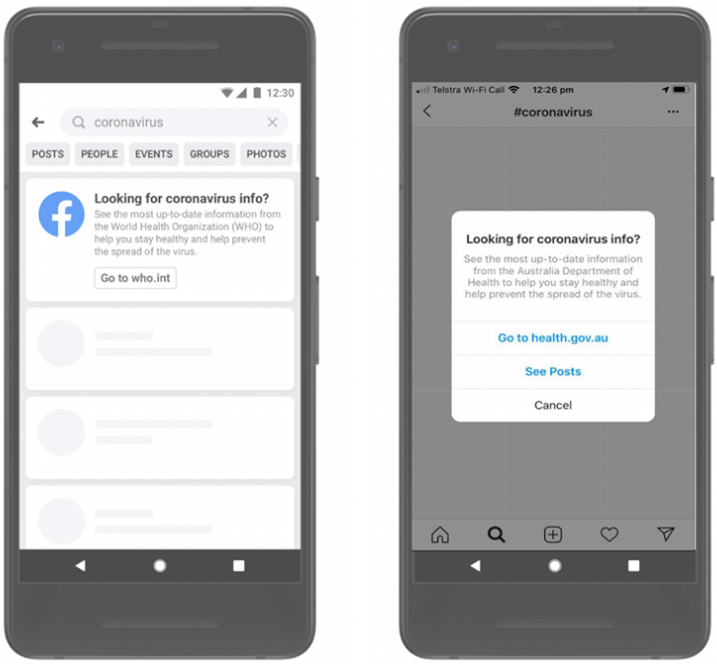
[39] Facebook, 'Stepping up the right against climate change', *Facebook Newsroom*, 14 September 2020, https://about.fb.com/news/2020/09/stepping-up-the-fight-against-climate-change/

- **Facebook gives significant amounts of ad credits to authoritative organisations, including the Australian Government and state and territory governments, to promote authoritative information.**

  Throughout 2020, we made a significant amount of ad credits freely available to the Australian Government and state and territory governments, to help extend the reach of their authoritative communications on our platforms. This was in addition to other measures to directly promote authoritative advice from governments on our platforms.

  Similarly, we have written to the Australian Government and offered to provide a second round of free advertising credits on our services in order to support the communications campaign about the COVID-19 vaccine rollout.

- **Facebook directs users to authoritative information when they search for high-priority topics on Facebook.** When people search for 'coronavirus' on our platforms, they are directed to the WHO, Coronavirus Information Centre or the Australian Health Department.

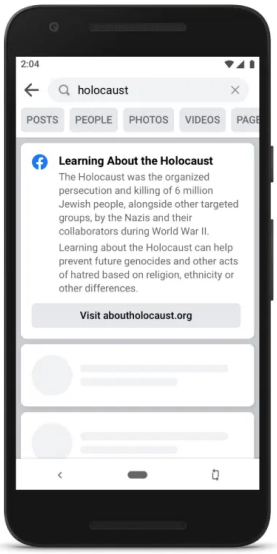


We have also used Search Redirect initiatives in relation to other types of harmful misinformation, for example, the QAnon conspiracy theory.

When someone searches for terms related to QAnon on Facebook and Instagram, we redirect them to credible resources from the Global Network on Extremism and Technology (GNET), the academic research network of the

Global Internet Forum to Counter Terrorism. Similarly, at one point in 2020, QAnon adherents took advantage of the #savethechildren hashtag to recruit and organise, and we have instituted a search prompt that promotes visits to the global charity Save the Children.

In January 2021, we also announced that we are now searching people to credible information about the Holocaust, when they search for terms associated with the Holocaust or Holocaust denial.[40]

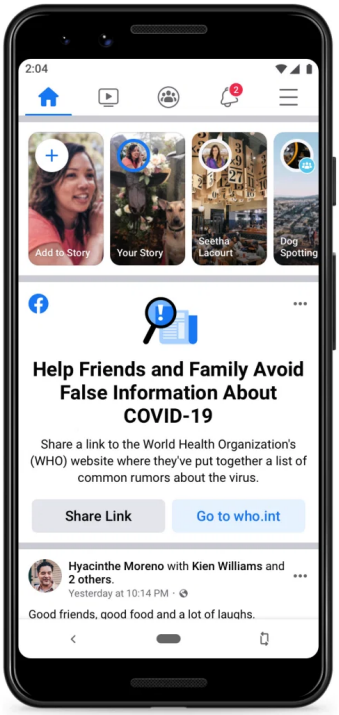


- **Facebook directs users to authoritative information once they have seen or shared COVID-19 related misinformation.** We have started showing messages in News Feed to people who have liked, reacted or commented on harmful misinformation about COVID-19 that we have since removed. These messages will connect people to COVID-19 myths debunked by the WHO[41] including ones we've removed from our platform for leading to imminent physical harm.

  If someone has shared a story that is later determined by fact-checkers to be false, we notify them that there is additional reporting on that piece of content.

[40] G Rosen, 'Connecting people to credible information about the Holocaust on Facebook', *Facebook Newsroom*, 21 January 2021, https://about.fb.com/news/2021/01/connecting-people-to-credible-information-about-the-holocaust-off-facebook/

[41] *See e.g.* https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters



- **Facebook will look for opportunities to continue to work with the Government on other ways to promote authoritative information.** We worked with the Digital Transformation Agency, Atlassian and service provider Turn.io to launch a chatbot on WhatsApp to help people easily access the latest information.

In early 2021, we wrote to the Government to propose further collaboration in order to provide authoritative information about the availability and locations of vaccines.

- **Facebook promotes public service announcements to our users to encourage them to be wary of potential misinformation.** Facebook has undertaken a number of large-scale public service announcement campaigns to encourage media literacy and to support users with identifying misinformation. We established the website fightcovidmisinfo.com, and undertook advertising in Australia - all to share six simple tips for users on how to spot and identify misinformation online.

  We intend to follow this up with further measures in Australia over the course of 2021.

Outcome 5
Users are better informed about the source of Political Advertising

Signatories will develop and implement policies that provide users with greater transparency about the source of Political Advertising carried on digital platforms.

- **Facebook requires all advertisers of political ads[42] to complete an ad authorisation, which includes verifying the advertiser's identity.** Since August 2020, we have required that any advertiser in Australia who is running political ads requires prior authorisation.[43] This includes verification of the advertiser's identity using official identity documentation in the country that they would like to advertise. We periodically check for accuracy and if provided information appears to be invalid or unavailable, the relevant ads will be taken down.

- **Facebook requires political ads to include a disclaimer disclosing who is paying for the ad.** To help journalists, civil society and the general community understand who is behind political ads, we require that any political ads also include a disclaimer that accurately reflects the organisation or person paying for the ads.

- **Facebook provides the Ad Library, a searchable archive of all political ads on our services in Australia, and will continue to add functionality to encourage scrutiny of political advertising.**

  The Ad Library is an industry-leading transparency initiative, which provides information to anyone about the ads on Facebook and Instagram, and who is behind them.

  The Ad Library has been available for some time, but we have been progressively adding different features.[44] Some of the key updates have been:

  - in April 2019, we updated the Ad Library to include all active ads any Page is running, along with more Page information such as creation

[42] We define political ads as advertisements: (1) made by, on behalf of, or about a candidate for public office, a political figure, a political party or advocates for the outcome of an election to public office; or (2) about any election, referendum or ballot initiative, including "go out and vote" or election campaigns. We recognise the definition of Political Advertising in the voluntary industry code for disinformation and misinformation is broader than Facebook's definition of "political ads", as it also encompasses ads that we refer to as "social issue ads".

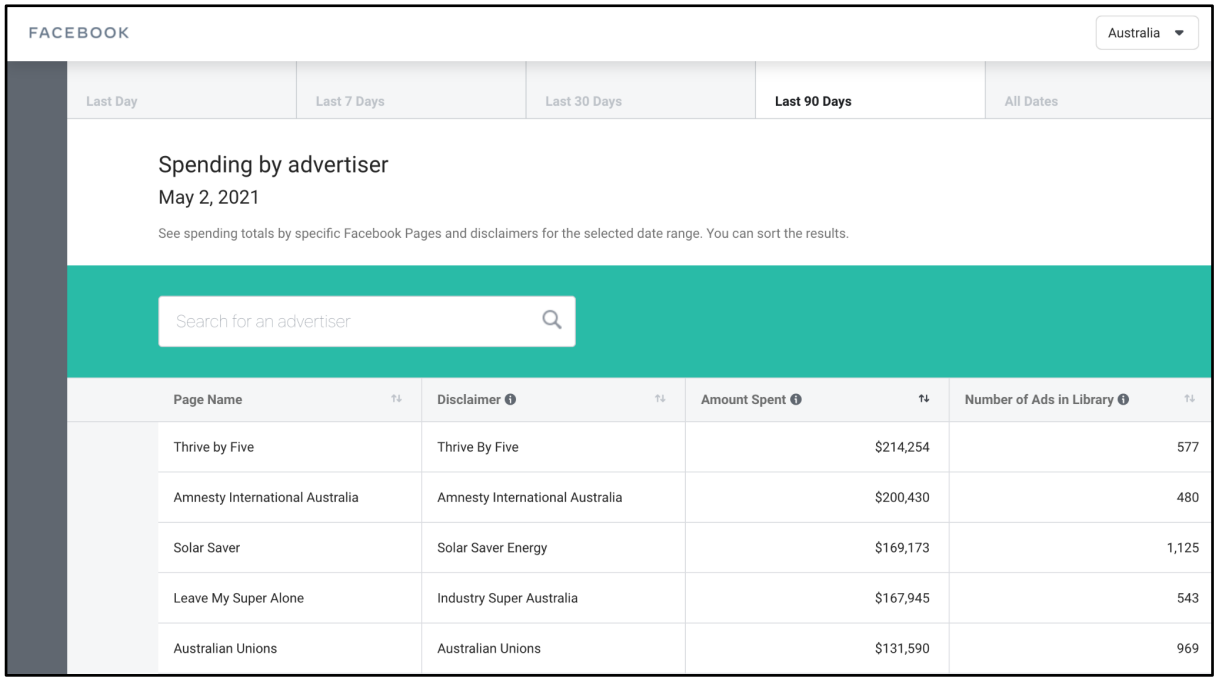
[43] Facebook, *Ads about Social Issues, Elections and Politics*, https://www.facebook.com/business/help/1838453822893854

[44] S Schiff, *Offering Greater Transparency for Social Issue, Electoral and Political Ads In More Countries*, https://about.fb.com/news/2019/06/offering-greater-transparency/; R Leathern, *Expanded Transparency and More Controls for Political Ads*, https://about.fb.com/news/2020/01/political-ads/

date, name changes, Page mergers and the primary country of people who manage Pages with large audiences.

- in August 2020, we instituted new requirements to run political ads on Facebook in Australia. We now verify the identity of every Page that runs political ads in Australia, and the Page admin must confirm they are currently living in Australia. The ads must also have a disclaimer with the name and entity that paid for the ads. And political ads are no longer just available while they are running; they are also stored in a searchable archive for up to seven years.
- We also now allow an Ad Library API which encourages greater scrutiny of advertisers and Facebook.

- **Facebook enables an Ad Library report that provides aggregated spend information about Pages undertaking political ads.** In August 2020, we began making available in Australia the Ad Library Report (available at facebook.com/ads/library/report) which allows for aggregated analysis of political ads on Facebook. It allows for easy comparison and analysis between advertisers.



- **Facebook will extend the policies and enforcement for political ads to social issue ads in 2021.** We will continue to progressively add functionality and strengthen enforcement in relation to ads on our services in Australia during 2021. We commit that we will bring the industry-leading levels of transparency and enforcement that we currently have for political ads to social issue ads.

Outcome 6
Signatories support the efforts of independent researchers to improve public understanding of Disinformation and Misinformation.

Signatories commit to support and encourage good faith independent efforts to research Disinformation and Misinformation both online and offline.

Facebook will continue to support research and events in relation to misinformation and media literacy. Facebook already supports a large amount of research and events to encourage policy understanding and debate about misinformation and media literacy. Some recent highlights include:

- Facebook provided funding (via the US-based National Association for Media Literacy Education) to support work done by academics from the Western Sydney University, Queensland University of Technology, and the University of Canberra to undertake Australia's first-ever nationwide adult media literacy survey. This was launched during a symposium in April 2021, held simultaneously across Sydney, Canberra and Brisbane to discuss misinformation and media literacy.
- We have invested US$2 million in a global round of funding for academic research on misinformation and polarisation. We announced the winners in August 2020, two of whom came from Australian universities.[45]
- We commissioned independent research by respected Australian academic Dr Andrea Carson to map government approaches to combating misinformation around the world, focussing on the Asia-Pacific region. The resulting report - Tackling Fake News - was launched in January 2021.
- We are working with Australian Associated Press to develop a media literacy initiative for Australians about the importance of fact-checking and how to recognise and avoid the spread of misinformation.
- Facebook was one of two major sponsors for the research work undertaken by First Draft and the University of Technology Sydney about the misinformation environment in Australia, to support the development of DIGI's disinformation and misinformation code.
- Facebook supported a First Draft event held in March 2021 to discuss health misinformation, where a digital panel was convened to discuss multiple facets of the problem.

[45] A Leavitt, K Grant, 'Announcing the winners of Facebook's request for proposals on misinformation and polarization', https://research.fb.com/blog/2020/08/announcing-the-winners-of-facebooks-request-for-proposals-on-misinformation-and-polarization/, 7 August 2020.

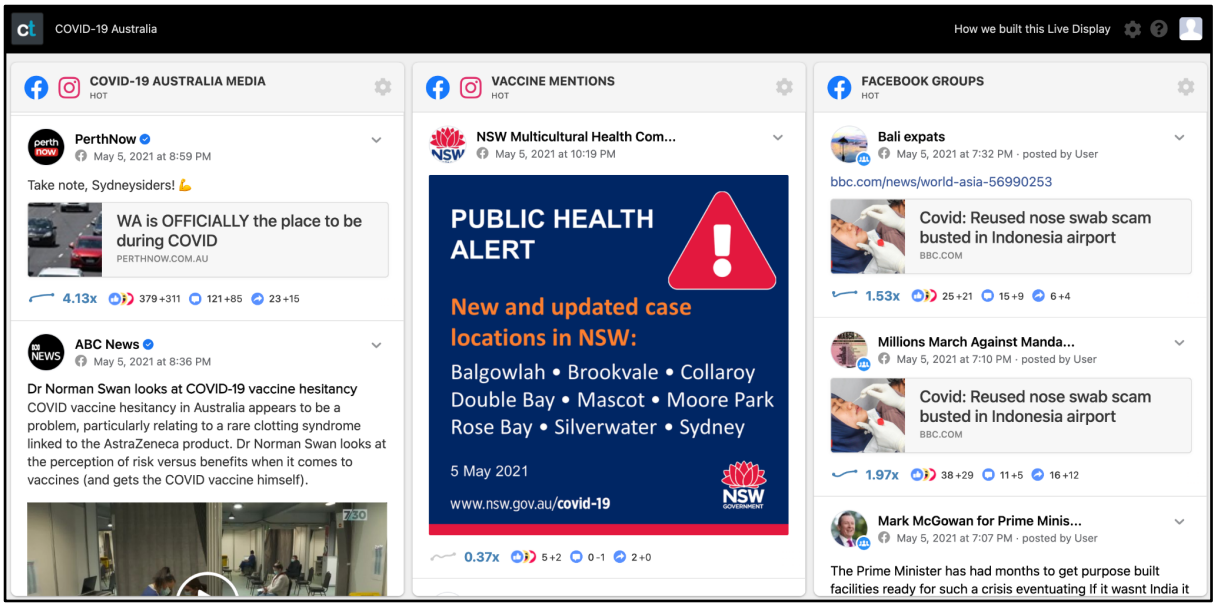
Facebook will continue to support research and events relating to misinformation and media literacy in 2021. In particular, we already have a partnership underway with misinformation and media literacy experts First Draft, with a number of deliverables expected for the rest of the year.

- **Facebook will continue to support research and events in relation to disinformation.** In addition to research and work on misinformation, Facebook has also been supporting work related to disinformation. We have sponsored Dr Jake Wallis to undertake a review of disinformation-for-hire, specifically targeting Australia and the Asia-Pacific region, and that research is expected to be completed and released in the coming months.

  We also continue to support the work of the Australian Strategic Policy Institute (ASPI), one of the world's leading experts in disinformation and malicious cyber activity, as a major sponsor.

- **Facebook provides a free CrowdTangle public live display on COVID-19 publicly available to allow anybody to analyse public content on our platforms.** While access to CrowdTangle is only to news publishers, third party fact-checking partners and approved academics, we recognised at the beginning of the COVID-19 pandemic that policymakers, advocates and the broader community would benefit from being able to understand the nature of COVID-19 discussions on our platform.

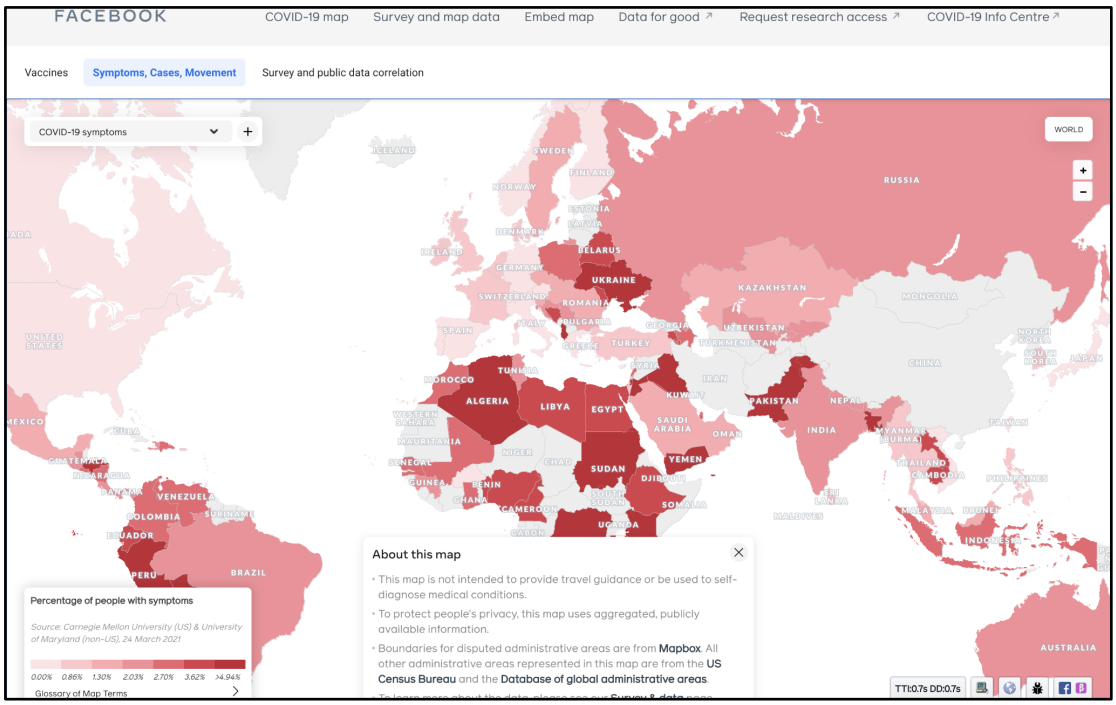
  To that end, we built a public live display that tracks conversations on Facebook and Instagram related to COVID-19. It is available at: https://apps.crowdtangle.com/covid19/boards/covid-19australia. This is essentially a 'live list' of COVID-19 content on our platform, which allows for the detection of misinformation that may be attracting a significant amount of attention. This helps to support the work of academics, experts, journalists and others in understanding the potential spread of misinformation.



- **Facebook collaborates with researchers to undertake surveys of our users to assess their views on topics such as vaccines and climate change.** We notably launched the COVID-19 Symptom Survey, one of the largest global data collection efforts related to COVID-19.[46] The World Symptom Survey is a partnership between Facebook and academic institutions. The survey is available in 56 languages. A representative sample of Facebook users is invited on a daily basis to report on symptoms, social distancing behavior, mental health issues, and financial constraints. Facebook provides weights to reduce nonresponse and coverage bias.

  Country and region-level statistics are published daily via public API and dashboards, and microdata are available for researchers via data use agreements. Over half a million responses are collected daily. This work has been done in partnership with Carnegie Mellon University and the University of Maryland.

[46] Facebook, *COVID-19 Symptom Survey*, https://dataforgood.fb.com/docs/covid-19-symptom-survey-request-for-data-access/.



- **Facebook provides data to researchers in a privacy-protective way via the Facebook Open Research and Transparency initiative.**

  Facebook Open Research and Transparency provides academics and independent researchers with tools and data to study Facebook's impact on the world. We accomplish this by developing Researcher Products with input from the academic community, including data access policies, analysis tools, data sets and APIs. Our goal is to facilitate privacy-protected data sharing, as well as the publication of independent, credible, and objective research for societal benefit and to advance platform accountability.[47]

  In addition to the researcher access outlined earlier this document, we also provide:

  - **the Researcher Platform.** The Facebook Open Research and Transparency Researcher Platform is a secure way for qualified users to access privacy-protected Facebook and Instagram data. It is built with validated privacy and security protections, such as data access controls, and has been penetration-tested by experts.

    The Researcher Platform runs a modified version of Jupyter, an open source tool that supports multiple standard statistical packages including, SQL, Python, and R, as well as a bridge to Facebook Graph APIs. Once a researcher applies and is approved to use the Researcher

[47] Facebook, *Facebook Research*, https://research.fb.com/data/

Platform, they gain access to 'virtual data clean rooms' where they can upload their own data and join it with Facebook data in a privacy-protected environment.

- **Researcher Pages API**. This API enables search across all public Pages on Facebook, including Posts and hashtags in Posts. The data includes Page name, description, country of Page owner and others.[48]

- **Researcher Analytics API.** This API includes a collection of API endpoints that helps academics identify trends on Facebook Pages and how they've evolved over time. Researchers can leverage these insights to focus on specific Pages that are of interest.[49] This API was announced only in March 2021.

[48] Facebook, *FORT Pages API*, https://developers.facebook.com/docs/fort-pages-api/

[49] T Lohman & K Jagadeesh, 'New analytics API for researchers studying Facebook Page data', *Facebook Research blog*, https://research.fb.com/blog/2021/03/new-analytics-api-for-researchers-studying-facebook-page-data/.

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

All Signatories will make and publish the annual report information in section 7

Facebook will continue to publish annual reports in Australia, such as these, to be transparent about the steps we are taking to combat disinformation and misinformation.

This will supplement the additional measures outlined above under Outcome 1.

As outlined above, Facebook has built a dedicated website that outlines our efforts to combat misinformation. https://www.facebook.com/combating-misinfo

Facebook makes available a detailed list of claims that we consider to violate our COVID-19 Misinformation & Harm policy.