

Meta response to the Australian Code of Practice on Disinformation and Misinformation

Reporting period: January - December 2023

Summary

As a founding member and signatory to the Australian Code of Practice on Disinformation and Misinformation (**the Code**), Meta provides our latest transparency report to publicise the measures that we take to combat disinformation and misinformation.

This latest update outlines Meta's approach to combat disinformation and misinformation in Australia during the 2023 calendar year. There are several key trends and changes that have occurred during the reporting period:

- **Global adversarial threats:** In 2023, we disrupted a number of Co-ordinated Inauthentic Behaviour (CIB) networks. More than half of these CIB networks targeted audiences outside of their countries of operation. We removed the majority of these networks before they were able to build authentic audiences. We identified four trends in CIB throughout 2023, specifically: an increase in China-based CIB disruptions, for-hire surveillance operations being behind CIB globally, abuse of domain name infrastructure and the Russian network 'Doppelgänger' trying to stay online.
- **Changed public health emergency declarations:** As the COVID-19 situation evolved, in mid-2022, we sought advice from Meta's Oversight Board specifically on whether we should change our approach to COVID-19-related misinformation. The Oversight Board issued an opinion finding that Meta should prepare measures for when the World Health Organisation lifted its public health emergency declaration, in order to protect freedom of expression and other human rights. In June 2023, we released our response to the Oversight Board's recommendations,¹ announcing that we would take a more tailored approach to our COVID-19 misinformation rules, in line with the Board's recommendations: in countries that have a COVID-19 public health emergency declaration, we will continue to remove content for violating our COVID-19 misinformation policies, given the risk of imminent physical harm. Globally, our COVID-19 misinformation rules are no longer in effect, as the global public health emergency declaration that triggered those rules has been lifted. Our commitments under the Code have therefore been

¹ Oversight Board, , ['Oversight Board publishes policy advisory opinion on the removal of COVID-19 misinformation'](#), April 2023.

adjusted to reflect this policy change.

- **Broader deployment of Generative AI:** Throughout 2023 and ongoing in 2024, there has been an increased awareness of the use of AI technology as part of content governance systems and also an increased focus on distribution, and use by businesses and consumers, of Generative AI technology. To respond to questions about responsible innovation, we began re-evaluating our policies and consulted with a range of stakeholders globally to identify if we needed to update and amend these to keep pace with the rapid generative AI advances. These consultations, as well as new recommendations by Meta's Oversight Board, led to a range of new changes specifically relating to how we handle manipulated media.

Looking ahead to 2024, Meta has opted into 40 commitments under the code and since the Code took effect in 2021, we have released three transparency reports outlining our efforts to meet these commitments.

In 2023, Meta's transparency report outlined 38 commitments to meet our obligations across both Facebook and Instagram. This report outlines the steps we took to meet these commitments across the reporting period of January to December 2023.² We have provided new case studies on the global trends in Coordinated Inauthentic Behaviour for 2023, as well as our work during the 2023 Aboriginal and Torres Strait Islander Voice Constitutional Referendum, where we introduced a number of specific measures to combat misinformation and disinformation.

We also continue to release up-to-date, Australia-specific data on our misinformation efforts to maintain transparency with the community and policymakers about the impact of these misinformation efforts. Between 1 January and 31 December 2023:

- We took action on over 9,700 pieces of content across Facebook and Instagram in Australia for violating our Misinformation policies.
- In addition to this, over 6,200 ads were removed in Australia for violating our Misinformation policy.
- We displayed warnings on over 9.2 million distinct pieces of content on Facebook, and over 510,000 on Instagram, in Australia (including reshares) based on articles written by our third-party fact-checking partners.
- We removed over 75,000 ads in Australia for not complying with our Social Issues, Elections and Politics (SIEP) ads policy.

In 2023, some further highlights of our work included:

- Implementing a suite of measures in the lead up to the 2023 Voice Referendum to proactively detect and remove content that breaches our policies, combat

² This reporting period has been specified in the guidelines provided by the independent reviewer, engaged by DIGI.

misinformation, harmful content and coordinated inauthentic behaviour, and promote civic participation.

An example of these efforts are the referendum day reminders we deployed to remind people to vote in the referendum.

- Our Civic Products reached around 12.7 million users on Facebook with around 13.6 million impressions.
 - For Instagram, around 6.55 million users were reached with around 6.57 million impressions.
- Investing in new research and analysis relating to information integrity surrounding the referendum debate and its outcomes, including in partnership with La Trobe University and RMIT CrossCheck.
- Sponsoring and hosting the Australian Media Literacy Alliance's Australian Media Literacy Summit, which brought together a range of journalists, academics, educators, librarians and other experts to discuss and learn about diverse ways to strengthen media literacy education in Australia.
- Supporting Western Sydney University's 2023 Young People and News longitudinal survey, which provides findings about the news attitudes, practices and experiences of young Australians aged between 8-16 years.
- Introducing new tools to provide users with more transparency around content ranking algorithms and recommendation systems, including updates to the Facebook and Instagram System Cards.
- We conducted significant research and collaborated with key industry partners across 2023 to produce a set of signals that will help us to label content that we believe has been made with AI. We believe providing transparency and additional context is now the better way to address manipulated media and avoid the risk of unnecessarily restricting freedom of speech, so we'll keep this content on our platforms so we can add labels and context.
- Preparing for and announcing a new AI Disclosure policy that allows advertisers to label Social Issues, Elections, and Politics (SIEP) Ads when they have been digitally created or altered so that viewers are aware. This policy went into effect in 2024.
- Sponsoring events in relation to promoting authoritative information, including a new training series called 'Connect, Alert, Inform', for emergency response organisations across Australia, New Zealand and the Pacific Islands. The training focused on helping to strengthen emergency communicators' skills in using Meta's services to build community and deliver critical disaster-related information. It included a module specifically developed and delivered by RMIT CrossCheck for

disaster communicators on combating, addressing and avoiding the amplification of misinformation surrounding a disaster or emergency situation.

In support of the Code and making meaningful developments in our transparency efforts, we have included two new metrics to this report. These are:

- The number of ads removed for violating our Misinformation policy from January to December 2023 in Australia.
- The number of ads removed for not complying with Meta's SIEP advertising policies,³ from January to December 2023 in Australia.

These steps are part of our extensive global efforts to combat misinformation and disinformation.

This report also outlines Meta's 40 commitments for the next reporting period. Meta will continue to document and report on relevant updates, changes and developments to our integrity practices related to disinformation and misinformation.

We look forward to continuing to work with Australian policymakers, civil society, academics and experts on steps to combat misinformation and disinformation in Australia over the next year.

³ Meta, '[Ads about Social Issues, Elections or Politics](#)', Meta Transparency Centre, 19 March 2024

List of Meta commitments under the Australian industry code on misinformation and disinformation for 2024

A copy of the 2023 commitments can be found in **Appendix A**.

The 40 commitments for 2024 are outlined below:

Outcome 1: Combatting misinformation / disinformation	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. <ul style="list-style-type: none">• Meta removes networks of accounts, Pages and Groups that violate our inauthentic behaviour policy, including disinformation that violates our policy on Inauthentic Behaviour.• Meta provides transparency about accounts, Pages and Groups removed under our Inauthentic Behaviour policy.• Meta partners with experts and organisations who assist in providing tips or further investigation about possible inauthentic behaviour on our services.• Meta removes misinformation that violates the misinformation section of our Community Standards.• Meta removes manipulated videos, also known as “deepfakes”, that violate our Manipulated Media policy.• Meta removes election-related misinformation that may constitute voter fraud or interference under our Misinformation, Coordinating Harm and Promoting Crime policies.• Meta removes election-related misinformation that may constitute voter suppression.• Meta removes fake accounts.• Meta allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.• Meta partners with third-party fact-checking organisations, globally and in Australia, to assess the accuracy of content on our services.• Meta applies a warning label to content found to be false, partly false, altered, or missing context by third-party fact-checking organisations.
--	--

- Meta reduces the distribution of content found to be false, partly false, or altered, by third-party fact-checking organisations.
- Meta proactively searches for content that matches content debunked by our fact-checking partners, to apply the same treatments.
- Meta takes action on Pages, Groups, accounts or websites found to repeatedly share misinformation.
- Meta continues to build safeguards into our GenAI features and models so that people can have a safer and more enjoyable experience.

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.

- Meta makes information available via a dedicated website that outlines our efforts to combat misinformation.

Outcome 1c. Users can report content and behaviours to Signatories that violate their policies under 5.10 through publicly available and accessible reporting tools.

- Meta makes on-platform reporting channels available to users for false information.
- Meta's Australian third-party fact-checking partners are also able to receive referrals from the public.

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

- Meta makes global transparency reports available regularly.
- Meta will supplement these reports with additional Australia-specific statistics, provided as part of this Annual Report process.

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

	<ul style="list-style-type: none"> • Meta will continue to provide ongoing transparency of our content ranking algorithms and give users more control over the content they see. • Meta takes steps to limit the possible distribution of misinformation via recommendations.
Outcome 2: Disrupt monetisation and advertising incentives	<ul style="list-style-type: none"> • Meta sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation. • Meta commits to defund the dissemination of misinformation.
Outcome 3: Combat inauthentic user behaviour	See items listed under Outcome 1.
Outcome 4: Empower consumers to be informed	<ul style="list-style-type: none"> • Meta provides contextual information around posts that users see from public Pages. • Meta provides consumers with on-platform tools to understand why they are seeing a particular post and to manage the content they see on Facebook and Instagram. • Meta reduces the distribution and recommendation of problematic and low quality content on Facebook and Instagram. • Meta will provide ongoing transparency in relation to AI-generated images and content. • Meta will look for opportunities to work with the Australian Government or local organisations to promote authoritative information and/or media literacy.
Outcome 5: Political advertising	<ul style="list-style-type: none"> • Meta requires all advertisers of political and social issue ads to complete an ad authorisation, which includes verifying the advertiser's identity. • Meta requires political and social issue ads to include a disclaimer disclosing who is paying for the ad. • Meta provides the Ad Library, a searchable archive of all political and social issue ads on our services in Australia.

	<ul style="list-style-type: none"> • Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political and social issue ads. • Meta will explore ways to ensure that there is appropriate ads transparency in relation to AI.
Outcome 6: Research	<ul style="list-style-type: none"> • Meta supports research and events in relation to misinformation and media literacy. • Meta supports research and events in relation to disinformation. • Meta collaborates with researchers to undertake surveys of our users to assess their views on relevant social issues. • Meta will make the Meta Content Library and API research tools available to third-party fact-checking partners and qualified users. • Meta will continue to support new areas of research relating to misinformation and disinformation in 2024.
Outcome 7: Annual reports	<ul style="list-style-type: none"> • Meta 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

Summary	1
List of Meta commitments under the Australian industry code on misinformation and disinformation for 2024	5
Table of contents	9
Reporting against 2023 commitments	10
Outcome 1a	10
Case study: Meta's response to global CIB attempts and global adversarial threats throughout 2023.	11
Outcome 1b	20
Outcome 1c	21
Outcome 1d	22
Outcome 1e	24
Outcome 2	26
Outcome 3	27
Outcome 4	28
Case study: Aboriginal and Torres Strait Islander Voice referendum	30
Outcome 5	35
Outcome 6	37
Outcome 7	40
Appendix A: 2023 specific commitments made by Meta under the industry code on misinformation and disinformation	41

Reporting against 2023 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

Our approach to misinformation and disinformation is consistent with that which was outlined in our 2021, 2022 and 2023 transparency reports.⁴ Below we outline developments from 2023:

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

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports. As mentioned in the baseline 2021 report, Meta's approach focuses on removing misinformation which could directly contribute to the risk of imminent physical harm, misinformation that could interfere with the functioning of political processes, certain highly deceptive manipulated media, as well as fake accounts (which are often vehicles for disinformation). It also focuses on reducing the spread of other misinformation, while promoting authoritative information. Further details of Meta's most recent efforts to remove coordinated inauthentic behaviour (CIB) are outlined in the commitment below.

- **Meta provides transparency about accounts, Pages and Groups removed under our Inauthentic Behaviour policy.**

We use a combination of policies, tools, expert teams and partnerships to detect and remove networks of inauthentic behaviour (IB) and CIB - both foreign and domestic.

We continue to report on our efforts to disrupt CIB through our Community Standards Enforcement Report and Quarterly Adversarial Threats report.⁵ As mentioned in our most recent report, Meta has removed over 200 covert influence

⁴ Meta, '[Meta's response to the Australian disinformation and misinformation industry code](#)', *Meta Australia blog*, May 2021.

⁵ Meta, [Community Standards Enforcement Report](#), and Meta, [Adversarial Threats report](#), Meta Transparency Centre.

operations between 2017-2022.⁶

In our 2023 transparency report, we reported that Meta had taken action on four instances of CIB operations that targeted Australians, since 2017. This number increased in 2023, to five instances as a China based network targeted Australia and other western countries. During the 2023 Aboriginal and Torres Strait Islander Voice Referendum, we did not see any evidence of coordinated inauthentic behaviour targeting Australia.

Below we provide a case study that looks at the adversarial threat trends globally across 2023.

Case study: Meta's response to global CIB attempts and global adversarial threats throughout 2023

A high level summary of CIB Networks disrupted in 2023:

- In Q1 we removed six separate covert influence operations for violating our policy against CIB. They originated in:
 - The United States & Venezuela
 - Iran
 - China (two separate networks originated in China)
 - Georgia
 - Burkina Faso & Togo
- In Q2 we removed three separate covert influence operations that violated our policy against CIB. They originated in:
 - Türkiye
 - Iran
 - China
- In Q3 we removed three separate covert influence operations that violated our policy against CIB. They originated in:
 - China (two separate networks originated in China)
 - Russia
- In Q4 we removed three separate covert influence operations that violated our policy against CIB. They originated in:
 - China
 - Myanmar
 - Ukraine

⁶ Meta, D Agranovich, '[Recapping our 2022 coordinated inauthentic behaviour enforcements](#)', *Meta Newsroom*, 15 December 2022.

More than half of these CIB networks targeted audiences outside of their countries of operation. We removed the majority of these networks before they were able to build authentic audiences.

Trends in coordinated inauthentic behaviour throughout 2023:

- **An increase in China based CIB disruptions:**
 - In Q2 2023, we removed a network of 7,704 Facebook accounts, 954 Pages, 15 Groups and 15 Instagram accounts for violating our policy against coordinated inauthentic behaviour. This network originated in China and targeted many regions around the world, including Australia, the United Kingdom, Taiwan, the United States, Japan, and global Chinese-speaking audiences.
 - We began this investigation after reviewing public reporting about off-platform activity that targeted a human rights NGO in late 2022. Following this lead, we were able to uncover a large and prolific covert influence operation which was active on more than 50 platforms and forums, including X (formerly Twitter), YouTube, TikTok, Reddit, Pinterest, Medium, Blogspot, LiveJournal, VKontakte, Vimeo, and dozens of smaller platforms and forums, as well as Facebook and Instagram.
 - On our platforms, this network was run by geographically dispersed operators across China who appear to have been centrally provisioned with internet access and content directions. Many of their accounts were detected and disabled by our automated systems. We assess that this likely led the people behind it to increasingly shift to posting its content on smaller platforms and then trying to amplify it on larger services in hopes to maintain persistence.
 - We have not found evidence of this network getting any substantial engagement among authentic communities on our services. This network typically posted positive commentary about China and its province Xinjiang and criticisms of the United States, Western foreign policies, and critics of the Chinese government including journalists and researchers.
 - Other China based networks identified in 2023 targeted India and Tibet, the United States, the EU, Taiwan, Sub-Saharan Africa, Japan, Central Asia and the Uyghur community around the world.

- Networks have utilised content in English and Chinese.
- **For-hire operations:**
 - Our 2023 reporting called out a trend that we continue to see, which is, for-hire surveillance organisations being behind covert influence operations globally, with many of the operations covered in our 4 quarterly reports being attributed to private entities. This included examples such as an IT company in China, a marketing firm in the United States and a political marketing consultancy in the Central African Republic, as well as several mercenary spyware companies operating in the EU despite EU privacy standards and regulations.⁷
- **Countering domain name abuse globally:**
 - Many threat actors continue to utilise domain name infrastructure in their malicious operations across the internet – from cyber espionage to covert influence campaigns and spyware firms. In 2023, we resolved a legal case against Freenom, a country code domain registry provider, whose domain names accounted for over half of all phishing attacks involving country code top-level domains (ccTLDs). This settlement resulted in Freenom announcing that it will exit the domain name business, including its operation of the country-code registries. While Freenom winds down its domain name business, it has agreed to treat Meta as a trusted notifier and it will also implement a block list to address future phishing, DNS abuse, and cybersquatting.⁸
- **The Russian network ‘Doppelganger’s’ persistence in trying to stay online:**
 - In 2022, we shared our threat research into the CIB network focused on supporting Russia’s invasion of Ukraine, dubbed ‘Doppelganger’, that operated across the internet, including running a large network of websites spoofing legitimate news outlets. In December 2022, we attributed it to two companies in Russia: Structura National Technology and Social Design Agency (Агентство Социального Проектирования). We banned these firms from our services. They were also later sanctioned by the EU.⁹ Our analysis from 2023 noticed several trends:

⁷ Meta, [Adversarial Threats report Q1 2023 and Q4 2023](#), Meta transparency Centre.

⁸ Meta, [Adversarial Threats report Q4 2023](#), Meta transparency Centre.

⁹ Meta, [Adversarial Threats report Q2 2023](#), Meta transparency Centre.

- Persistence: We assessed this campaign to be the largest and most aggressively persistent covert influence operation from Russia that has been seen since 2017. Since our initial disruption and continuous scrutiny by platforms and researchers, Doppelganger continued to create new domains in an attempt to evade detection. Given the nature of this operation and the type of entities behind it, this is expected behaviour across our industry with any CIB network we each take down.
- In addition to ongoing detection by our automated systems, our team has been monitoring and taking action against these recidivist attempts, and sharing findings with our peers and with the public. In total (as of Q3 2023), we've blocked over 2,000 of the operation's domains from being shared on our platform. We also blocked tens of thousands of attempts to run fake accounts and Pages on Meta's platforms.
- High input – low output: This operation stands out for the sheer wastefulness of its large-scale efforts. We expect Doppelganger to keep at it with its “smash-and-grab” approach by throwing a large amount of resources – even if it leads to a very high detection rate and loss of assets, as we described in September 2022 and in our last report.
- Expanding targeting, yet single mission: With Doppelganger focusing on weakening support for Ukraine against the Russia invasion, this operation appears to be trying to pick off some of Ukraine's international allies over time. Judging by the origin of the organisations that this operation spoofed, among other factors, this Russian campaign has expanded beyond targeting France, Germany and Ukraine itself for the first 8+ months to include the US and Israel earlier this year. While the exact reasoning behind this expansion is unknown, it likely reflects the fluid tasking of this operation (by its clients) and its single-minded mission.

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

Meta continues to maintain close relationships with experts and organisations around the world so we can share threat findings and adapt our enforcement.

- **Meta removes misinformation that violates the misinformation section of our Community Standards, and will review its Misinformation and Harm Policy, in line with recommendations from the Oversight Board.**

Between 1 January and 31 December 2023, we took action on over 9,700 pieces of content across Facebook and Instagram in Australia for violating our Misinformation policies. This is compared to 91,000 pieces of content for the 2022 reporting period and 180,000 (globally) in 2021.

The figure for 2023 above reflects that during this reporting period – as mentioned above – the Oversight Board provided a policy advisory opinion on the removal of COVID-19 misinformation, to which Meta subsequently responded. The Board prepared its advice in response to Meta’s request on whether measures to address dangerous COVID-19 misinformation, introduced in extraordinary circumstances at the onset of the pandemic, should remain in place as many, though not all, countries around the world were seeking to return to more normal life. As a result, in June 2023, Meta updated our Misinformation and Harm policy to reflect the Oversight Board’s guidance. We now only remove this content in countries with an active COVID-19 public health emergency declaration. This change has impacted our enforcement metrics on removals for this reporting period, but does not change our overall approach to fact-checking. In general also, changes to this metric are an expected part of fluctuating content trends online.

- **Meta removes manipulated videos, also known as “deepfakes”, that violate our Manipulated Media policy.**

Throughout 2023, Meta continued to enforce its policy on manipulated videos, as outlined in our 2021, 2022 and 2023 transparency reports.

However, in the first half of 2023, we began reevaluating our policies to see if we needed a new approach to keep pace with rapid advances in generative AI technologies and usage. We completed consultations with over 120 stakeholders in 34 countries in every major region of the world. Overall, we heard broad support for labelling AI-generated content and strong support for a more prominent label in high-risk scenarios. Many stakeholders were receptive to the concept of people self-disclosing content as AI-generated.¹⁰

A majority of stakeholders agreed that removal should be limited to only the highest risk scenarios where content can be tied to harm, since generative AI is becoming a mainstream tool for creative expression. This aligns with the principles behind our Community Standards – that people should be free to express themselves while also remaining safe on our services.

¹⁰ Meta, ‘[Our approach to labelling AI generated content and manipulated media](#)’, Meta Newsroom, April 5 2024.

We also conducted public opinion research with more than 23,000 respondents in 13 countries and asked people how social media companies, such as Meta, should approach AI-generated content on their platforms. A large majority (82%) favour warning labels for AI-generated content that depicts people saying things they did not say.¹¹

Additionally, the Oversight Board noted that their recommendations were informed by consultations with civil-society organisations, academics, inter-governmental organisations and other experts.

We conducted this research throughout 2023 and starting from May 2024, we will begin labelling a wider range of video, audio and image content as “Made with AI” when we detect industry standard AI image indicators or when people disclose that they’re uploading AI-generated content.¹²

- **Meta removes election-related misinformation that may constitute voter fraud or interference under our Coordinating Harm and Promoting Crime policy.**

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

We continue to maintain our approach to remove content that may cause voter suppression or misrepresentation of election related information as outlined in our 2021, 2022 and 2023 transparency reports.

- **Meta removes fake accounts.**

Meta continues to enforce its policy on fake accounts. We do not allow fake accounts on Facebook and Instagram, as they can be vehicles for a range of harmful content and behaviour.

Our ability to detect and remove fake accounts has been improving over the years, and there has been a general decline in the volume of fake accounts found on the platform since 2019.

From January to December 2023, we detected and removed 2.6 billion fake accounts on Facebook, on average we proactively detected and removed 98.95% of these accounts before they were reported to us.¹³ These are often caught within minutes of registration.

¹¹ Meta, [‘Our approach to labelling AI generated content and manipulated media’](#), Meta Newsroom, April 5 2024.

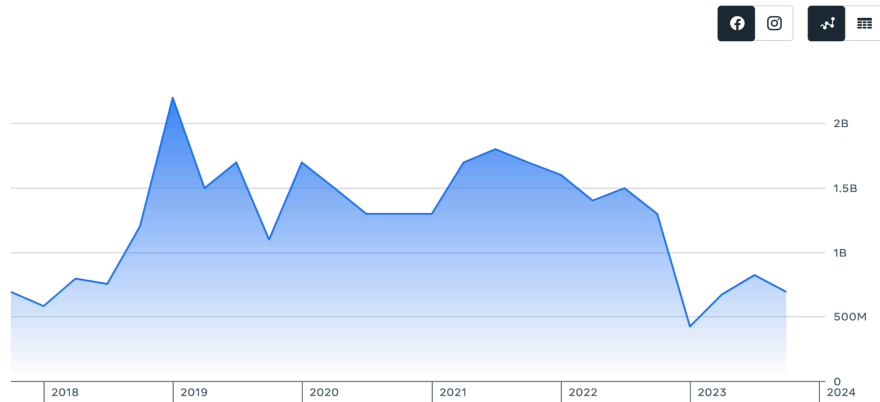
¹² Meta, [‘Our approach to labelling AI generated content and manipulated media’](#), Meta Newsroom, April 5 2024.

¹³ Meta, [‘Community Standards Enforcement Report’](#), Meta Transparency Centre.

of fake accounts we've taken action on (Oct 2017 - Dec 2023)¹⁴

ACCOUNTS ACTIONED

How many fake accounts did we take action on?



- **Meta allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.**

As mentioned in the 2021, 2022 and 2023 transparency report, Meta has voluntarily established the independent Oversight Board to ensure greater accountability for our content governance on Facebook and Instagram.

In our last report, we noted that the Oversight Board issued 36 case decisions and 176 recommendations to Meta for future improvements.

Between January and December 2023, the Oversight Board issued 53 decisions and 66 recommendations.¹⁵

The Oversight Board also investigated two cases relating to the Australian Electoral Commission's (AEC) voting rules around the 2023 Aboriginal and Torres Strait Islander Voice Referendum.¹⁶ The cases involved two separate Facebook posts containing the same screenshot of information posted on X by the AEC ahead of the Referendum. The information shown included the message that: 'If someone votes at two different polling places within their electorate, and places their formal vote in the ballot box at each polling place, their vote is counted.' The posts were removed by Meta for violating the rule in our Coordinating Harm and Promoting Crime Community Standard that prohibits content calling for illegal

¹⁴ Note - graph represents number of fake accounts removed globally. Australia-specific statistics are not available.

¹⁵ Oversight Board, '[H2 2023 Transparency Report](#)', and, Oversight Board, '[Q4 2023 Transparency report](#)'.

¹⁶ Oversight Board, '[Oversight Board announces cases involving the Australian Electoral Commission's voting rules](#)', Oversight Board News, February 2024.

participation in a voting process. Both users had appealed the removal of these posts.

On 9 May 2024, the Oversight Board published its decision on the two cases, in which it upheld Meta's decision to remove the posts, as the two users' 'calls for others to engage in illegal behaviour impacted the political rights of people living in Australia' and that Meta 'was correct to protect democratic processes by preventing voter fraud attempts from circulating on its platforms, given the frequent claims that the Voice Referendum was rigged'.¹⁷

- **Meta partners with third-party fact-checking organisations, globally and in Australia, to assess the accuracy of content on our services.**

Meta has built a large global fact-checking network. As of December 2023, Meta partners with over 90 fact-checking partners covering more than 60 languages globally.¹⁸

Australians benefit from Meta's international approach to fact-checking - an Australian user will see a warning label on content that has been fact-checked by a third-party fact-checking partner. Content found to be false by our international fact-checking partners (including our Australian fact-checkers listed below) will be demoted in an Australian user's Feed, and will not be recommended, meaning there is less chance of them seeing it.

In Australia, we partner with three third-party fact-checkers: Australian Associated Press (AAP), Agence France Presse (AFP) and RMIT FactLab.¹⁹ In preparation for the 2023 Aboriginal and Torres Strait Islander Voice Referendum, we provided a one-off funding boost to AAP and AFP, so that they could increase their capacity in the lead up to the referendum.²⁰

- **Meta applies a warning label to content found to be false by third-party fact-checking organisations.**

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

¹⁷ Oversight Board, '[Oversight Board upholds Meta's decisions in Australian Electoral Commission voting rules cases](#)', May 2024.

¹⁸ Meta, '[Where we have fact-checking](#)', *Meta for Media*.

¹⁹ *Note:* From 28 August 2023, RMIT Factlab was suspended from our third-party fact checking program pending the International Fact-Checking Network's decision on RMIT FactLab's expired certification. Following the IFCN's review and RMIT FactLab's recertification on 6 Nov 2023, Meta has since reinstated RMIT FactLab to its third-party fact-checking program.

²⁰ Meta, '[How Meta is preparing for the Voice to Parliament Referendum](#)', *Medium*, 9 July 2023.

The focus of Meta's fact-checking program is to identify and address viral misinformation, particularly clear hoaxes that have no basis in fact. Fact-checking partners prioritise probably false claims that are timely, trending and consequential. Meta applies a warning label to content found to be misinformation by the fact-checking organisations. Once fact-checkers have determined that a piece of content contains misinformation, Meta uses technology to identify near-identical versions of that content across Facebook, Instagram and Threads. Fact-check labels are only applied to near-identical versions of content that has already been rated; this means that labels are not applied to content that makes a similar claim but is differently expressed.²¹

Between 1 January and 31 December 2023, we displayed warnings on over 9.2 million distinct pieces of content on Facebook, and over 510,000 on Instagram, in Australia (including reshares) based on articles written by our third-party fact-checking partners. This compares with 9 million on Facebook in Australia for the 2022 reporting period, and a global number for 2021 on Facebook which was 190 million.

- **Meta reduces the distribution of content found to be false by third-party fact-checking organisations.**

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency report.

- **Meta proactively searches for content that matches content debunked by our fact-checking partners, to apply the same treatments.**

Meta uses our technology to detect content that is the same or near-identical versions to content that has been fact-checked by our fact-checking partners. Generally, we do not apply treatments such as warning labels to content that makes a similar claim rated by fact-checkers, if the content is not identical. This is because small differences in how a claim is phrased might change whether it is true or false.

Between January – December 2023, we displayed warnings on over 9.2 million distinct pieces of content on Facebook (including re-shares) and 510,000 on Instagram. This compares with 9 million on Facebook in Australia for the 2022

²¹ Following the submission and publication of Meta's 2022 transparency report in May 2023, a complaint was submitted by Reset Australia to the Independent Complaints Sub-Committee (under the Australian Code of Practice on Disinformation and Misinformation). The Committee has dismissed the complaint on the grounds that Reset Australia produced no convincing evidence that Meta's transparency report contained false statements (additional information on the complaint can be found [here](#)). To provide further clarity and transparency, Meta has updated the language in this report to ensure additional clarity regarding our third-party fact-checking program moving forward.

reporting period, and a global number for 2021 on Facebook which was 190 million.

- **Meta takes action on Pages, Groups, accounts, or websites found to repeatedly share misinformation.** We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

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.

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

As outlined in our previous transparency reports, during the pandemic, Meta made available a detailed list of claims that violated our COVID-19 Misinformation & Harm Policy. These claims were available in our help centre throughout 2023.²²

However, Meta's policies in relation to COVID-19 misinformation changed in June 2023 following advice and consultation from Meta's independent oversight board.²³ Our COVID-19 misinformation rules are no longer in effect globally as the global public health emergency declaration that triggered those rules has been lifted. During the pandemic, we removed content making one of 80 false claims about COVID-19 which public health authorities concluded were likely to directly contribute to the risk of imminent physical harm. We now only remove this content in countries with an active COVID-19 public health emergency declaration. This change has impacted our enforcement metrics for this policy globally.

- **Meta makes information available via a dedicated website that outlines our efforts to combat misinformation.**

²² Meta, '[COVID 19 and Vaccine Policy Updates and protections](#)', Facebook Help Centre.

²³ Oversight Board, '[Removal of COVID-19 Misinformation](#)', Oversight Board Decisions.

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency report. Meta's policies on misinformation can be found in our Transparency Centre.²⁴

Outcome 1c

Users can report content and behaviours to Signatories that violate 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.

- Meta makes on-platform reporting channels available to users for false information.

Meta continues to make reporting channels available to users on both Facebook and Instagram. As outlined in our 2021, 2022 and 2023 transparency reports, users can follow the following steps to report misinformation:

- Facebook Help Centre - <https://www.facebook.com/help/572838089565953>
- Instagram Help Centre - <https://www.facebook.com/help/instagram/2442045389198631>

Our Australian third-party fact-checking partners are also able to receive referrals from the public using the channels below:

- Agence France Presse: <https://factcheck.afp.com/contact>
- Australian Associated Press: <https://www.aap.com.au/make-a-submission/>
- RMIT FactLab: <https://www.rmit.edu.au/about/schools-colleges/media-and-communication/industry/factlab/debunking-misinformation>.

²⁴ Meta, '[Community Standards - Misinformation](#)', Meta Transparency Centre.

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.

- **Meta makes global transparency reports available regularly.**

Meta continues to make aggregated reports publicly available on our misinformation and disinformation efforts to provide oversight of our work. A range of compliance, regulatory and proactive transparency reports can be found in our online Transparency Centre.²⁵

As outlined in our 2021, 2022 and 2023 transparency reports, each quarter, we report on metrics for preventing and taking action on content that goes against our Community Standards. This is outlined in Meta's Community Standards Enforcement Report (CSER).²⁶

In 2022, Meta published the results of an independent audit, conducted by Ernst & Young (EY), into Meta's CSER reporting. EY found the calculation of the metrics in our 2021 fourth quarter Community Standards Enforcement Report were fairly stated, and our internal controls are suitably designed and operating effectively.²⁷ We continued publishing our CSER report throughout 2023.

- **Meta will supplement these reports with additional Australia-specific statistics, provided as part of this Annual Report process.**

While country-specific statistics should be interpreted with caution and have limitations in understanding misinformation and disinformation, Meta has again provided Australia-specific statistics in the spirit of transparency of our efforts.

²⁵ Meta, '[Meta Transparency Reports](#)', Meta Transparency Centre.

²⁶ Meta, '[Community Standards Enforcement Report](#)', Meta Transparency Centre.

²⁷ Meta, '[Community Standards Enforcement Report](#)', Meta Transparency Centre. The results of the EY assessment can be found at <https://about.fb.com/wp-content/uploads/2022/05/EY-CSER-Independent-Assessment-Q4-2021.pdf>.

Several Australia specific metrics are included throughout this report including:

- The number of pieces of content removed across Facebook & Instagram in Australia for violating our Misinformation policies.
- The number of warnings displayed on distinct pieces of content on Facebook and Instagram in Australia (including reshares) based on articles written by our third-party fact-checking partners.
- The number of ads removed on Facebook and Instagram combined for violating our Misinformation policy.
- Specific metrics relevant to our media literacy campaigns conducted for the Aboriginal and Torres Strait Islander Voice referendum.

In addition to this, while we were unable to report some metrics for this reporting cycle, we are working to implement changes which will allow us to report on the reshare friction metric in Australia in future reports.

In relation to enforcement regarding Misinformation at an account level. As noted above, Meta's policies to tackle false claims about COVID-19 which could directly contribute to the risk of imminent physical harm changed in June 2023 following Meta's independent Oversight Board's advice.²⁸ We now only remove this content in countries with an active COVID-19 public health emergency declaration. This change has impacted our enforcement metrics on removals for this reporting period, but does not change our overall approach to fact-checking. These changes are an expected part of fluctuating content trends online.

²⁸ Oversight Board, '[Oversight Board publishes policy advisory opinion on the removal of COVID 19 Misinformation](#)', Oversight Board News, 20 April 2023.

Outcome 1e.

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

Signatories that provide services (other than search engines) whose primary purpose is to disseminate information to the public and which use recommender systems, commit to :

- A. make information available to end-users about how they work to prioritise information that end-users may access on these services; and**
- B. provide end-users with options that relate to content suggested by recommender systems that are appropriate to the service.**

- **Meta will continue to provide greater transparency of our content ranking algorithms and give users more control over the content they see.**

In June 2023, we launched Facebook System Cards²⁹ and Instagram System Cards³⁰ to help people to understand how AI shapes their product experiences and provides insights into how ranking systems dynamically work to deliver a personalised experience on Facebook and Instagram. These cards are available on Meta's Transparency Centre, and clicking into a card will bring up more information regarding how that specific AI system works, for example on Facebook Feed. These system cards were written in a way that can be understood by experts and non-experts alike.³¹

In addition to the detailed system cards, in June 2023, Meta also expanded the tools available to users to help them control the content they see and personalise their experience.³²

We also updated the 'Why am I seeing this ad' tool to provide more transparency about how activity both on and off of Facebook or Instagram inform machine learning models that Meta uses to shape the content that users see.³³ We also extended the 'Why am seeing this' tool for content from Feed, to Instagram Reels and Explore.

²⁹ Meta, '[Our approach to explaining ranking](#)', Meta Transparency Centre, 31 December 2023 and Meta, '[Building Generative AI features responsibly](#)', 27 September 2023.

³⁰ Meta, '[Our approach to explaining ranking](#)', Meta Transparency Centre, 31 December 2023 and Meta, '[Building Generative AI features responsibly](#)', 27 September 2023.

³¹ Meta, '[Our approach to explaining ranking](#)', Meta Transparency Centre, 31 December 2023 and Meta, '[Building Generative AI features responsibly](#)', 27 September 2023.

³² Meta, '[How AI Influences What You See on Facebook and Instagram](#)', Meta Newsroom, 29 June 2023

³³ Meta, '[Increasing our Ads Transparency](#)', Meta Newsroom, 14 February 2023.

We have also created centralised places on Facebook and Instagram where users can customise controls that influence the content they see on each app. Specifically:

- Users can visit their Feed Preferences on Facebook through the three-dot menu on relevant posts, as well as through Settings. For Instagram, users can visit the Suggested Content Control Center through the three-dot menu on relevant posts, as well as through Settings.
 - We also made the “Show more, Show less” feature,³⁴ which is available on all posts for Facebook and Instagram Feed and Reels, as well as Facebook Video, even more prominent via the three-dot-menu.
 - If a user does not want an algorithmically-ranked Feed, they can use the Feeds tab on Facebook. There is also an option to add people to the Favourites list on Facebook. Likewise, users on Instagram can change to a chronological feed, Following, based on the accounts they follow.
 - The Reduce tool allows users to adjust the degree to which we demote problematic or low-quality content in their Feed (our Content Distribution Guidelines outline some of the most significant reasons why problematic or low-quality content may receive reduced distribution in Feed).
- **Meta takes steps to limit the possible distribution of misinformation via recommendations.**

Meta takes steps to limit the possible distribution of misinformation via recommendations. Facebook and Instagram content that has been debunked³⁵ by non-partisan, third-party fact-checking organisations who partner with Meta, and have been certified by the International Fact-Checking Network (IFCN) is removed from recommendations.³⁶

Pages, Groups, Profiles, websites, and Instagram accounts that repeatedly share content rated False or Altered will be put under some restrictions for a given time period. This includes removing them from the recommendations we show people, reducing their distribution, removing their ability to monetise and advertise and removing their ability to register as a news Page.³⁷

³⁴ Meta, [‘The new AI-powered feature designed to improve Feed for everyone’](#), Meta Blog, 5 October 2022.

³⁵ Meta, [‘About fact-checking on Facebook, Instagram and Threads’](#), Meta Business Help Centre.

³⁶ Meta, [‘fact-checked Misinformation’](#), Meta Transparency Centre.

³⁷ Meta, [‘Penalties for sharing fact-checked content’](#), Meta Transparency Centre.

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.

- **Meta sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation.**

We have maintained the policy approach outlined in our 2021, 2022 and 2023 transparency reports.

We removed over 6,200 ads for violating our Misinformation policy from January to December 2023 in Australia. We also removed over 75,000 ads for not complying with SIEP ads policies³⁸ from January to December 2023 in Australia. This report is the first instance in which we've shared this data.

³⁸ Meta, '[Ads about Social Issues, Elections or Politics](#)', Meta Transparency Centre, 19 March 2024.

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.

Please see Outcome 1a for the actions Meta takes against inauthentic user behaviours.

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.

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

With respect to these commitments, we have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports in terms of the contextual information around posts that users see from public Pages.

- **Meta provides a Climate Science Information Centre in Australia to connect users to authoritative information from leading climate organisations.**

Our platforms extend access to information and empower users to take action through tools like the Climate Info Finder Tool and the Climate Science Information Center, which launched in Australia in November 2021.³⁹ There are more than 18 million followers of the Center globally.⁴⁰

- **Meta uses in-product prompts to direct Australians to authoritative information on key topics.**

Consistent with previous reports under this code, we have launched in-app features that direct users to authoritative election information in Australia.

Ahead of the 2023 Aboriginal and Torres Strait Islander Voice Referendum, we developed reminder prompts to raise awareness about the referendum and link users to authoritative information.

More information is provided about these prompts in the Case Study: Aboriginal and Torres Strait Islander Voice Referendum in this report below.

³⁹ Meta, '[Our commitment to combatting climate change](#)', *Meta Newsroom*, 1 November 2021.

⁴⁰ Meta, '[Our approach to climate content](#)', 4 November 2022.

- **Meta promotes authoritative information by providing significant support to organisations such as the Australian, state and territory governments and First Nations health organisations to promote authoritative health information.**
- **Meta directs users to authoritative information when they search for high-priority topics on Facebook and Instagram.**
- **Meta directs users to authoritative information once they have seen or shared COVID-19 related misinformation.**

In response to **the three commitments above**, which focus on COVID-19-related information; as noted above, in June 2023 (in response to recommendations by Meta's Oversight Board), we announced that our global COVID-19 misinformation rules would be altered as the global public health emergency declaration that triggered those rules had been lifted. We now only remove this content in countries with an active COVID-19 public health emergency declaration. In light of this policy change, we have shifted our time and resources to respond to new and emerging priorities, such as content governance relating to AI technology.

- **Meta will look for opportunities to continue to work with the Government on other ways to promote authoritative information.**

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

As an example of this, in July 2023, Meta facilitated a new training series called 'Connect, Alert, Inform', for emergency response organisations across Australia, New Zealand and the Pacific Islands. The training focused on helping to strengthen emergency communicators' skills in using Meta's services to build community and deliver critical disaster-related information.

The virtual training curriculum was developed in consultation with a disaster communications academic from the University of Technology Sydney, and in partnership with Emergency Management and Public Affairs (EMPA) and RMIT CrossCheck. It comprised sessions specifically tailored for emergency responders across a range of topics, including Meta's crisis management tools; planning social media content; page moderation and account security; combating disaster related mis and dis-information; and social media advertising.

The training also included a separate simulation session, involving a fictional disaster scenario, where participants were encouraged to apply the learnings from the plenary training and consider how to best utilise social media as part of their emergency preparedness and response toolkit, and respond to challenges such as disaster-related misinformation and scams.

A diverse range of emergency communicators participated in the training, representing a mix of primarily government emergency response agencies, as well as law enforcement and humanitarian and disaster non-profit organisations from across the region.

- **Meta promotes public service announcements to our users to encourage them to be wary of potential misinformation.**

In terms of connecting Australians with authoritative information about key topics, in 2023, this was focused on the Aboriginal and Torres Strait Islander Voice Referendum, climate science information and COVID-19.

Ahead of the 2023 Aboriginal and Torres Strait Islander Voice Referendum, we developed reminder prompts to raise awareness about the referendum and link users to authoritative information.

You can find more information about these prompts in the Case Study: Aboriginal and Torres Strait Islander Voice Referendum.

Case study: Aboriginal and Torres Strait Islander Voice Referendum

In the lead up to Australia's Aboriginal and Torres Strait Islander Voice Referendum in October 2023, Meta developed a comprehensive strategy that focussed on proactively detecting and removing content that breaches our policies, combating misinformation, harmful content and CIB, and promoting civic participation.

Meta has been involved in more than 200 elections around the world since 2016, and we've learned key lessons from each one on combating misinformation, election interference and promoting civic participation. Alongside our routine integrity measures which we implemented for previous elections, we also rolled out additional measures to promote safety and integrity across our platforms in advance of the referendum.

This case study provides a comprehensive outline of our referendum integrity measures and their impact in Australia.

In the 2023 referendum campaign period, we developed referendum day reminders on Facebook and Instagram to encourage people to vote. Examples of these prompts for Facebook and Instagram, and stickers for Instagram can be found at the end of this case study.

These prompts reached a large number of Australians:

- Our Civic Products reached around 12.7 million users on Facebook with around 13.6 million impressions.
- For Instagram, around 6.55 million users were reached with around 6.57 million impressions.

We initiated a number of new programs of work in the lead up to the referendum in relation to misinformation and disinformation.⁴¹ This work includes:

- **Combating Misinformation and Foreign Interference Expanding Capacity for Meta's Australian Fact-Checkers:** We know the importance of ensuring Australians had access to reliable information about the Aboriginal and Torres Strait Islander Voice referendum in Australia. Therefore, we provided a one-off funding boost to AAP and AFP, to increase their capacity in the lead up to the referendum. Our fact-checkers are independent and work to reduce the spread of misinformation across Meta's services. When they rate something as false, we significantly reduce its distribution so fewer people see it. We also notify people who try to share something rated as false and add a warning label with a link to a debunking article.
- **Empowering People to Identify False News:** Since we know it's not enough to just limit or remove harmful or misleading misinformation that people see, we launched a new media literacy campaign with Australian Associated Press,⁴² building on our "Check The Facts"⁴³ campaign which ran ahead of the 2022 Federal Election in October 2021 and early 2022. This shared tips and advice with people so that they could make informed decisions about what they read, trust and share. Further information about this campaign can be found below in the media literacy commitments under Outcome 6.
- **Combatting Influence Operations:** We have specialised global teams which worked during the referendum to identify and take action against threats to the elections and referendums, including signs of coordinated inauthentic behaviour across our apps. We also coordinated with the Government's election integrity assurance taskforce and security agencies in the lead up to the referendum. We improved our AI so that we could more effectively detect and block fake accounts, which are often behind this activity.
- **Supporting academic analysis:** We supported analysis relating to information integrity surrounding the referendum debate and its outcomes, including RMIT CrossCheck's work to promote accurate and corrective information on the

⁴¹ Meta, '[How Meta is preparing for the Voice to Parliament Referendum](#)', *Medium*, 9 July 2023.

⁴² AAP, '[AAP FactCheck and Meta media literacy collab](#)', AAP Press release, 28 August 2023.

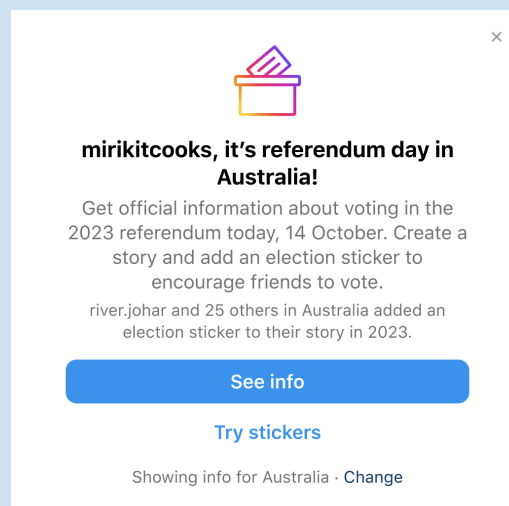
⁴³ AAP, '[Check the Facts](#)', AAP Resources.

referendum debate and boost media literacy. We also supported La Trobe University's report 'Influencers and Messages: Analysing the 2023 Voice to Parliament Referendum Campaign',⁴⁴ which examined the main topics of the debate, key actors and campaign strategies for Yes and No, and the prevalence and influence of misinformation and disinformation.

Referendum day reminders used in the build up to and during the referendum on Facebook:



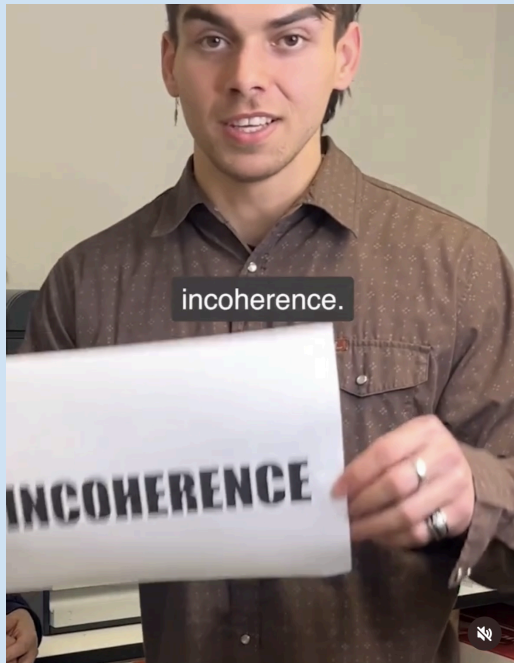
Referendum day reminders and stickers used in the build up to and during the referendum on Instagram:



⁴⁴ Andrea Carson, Rebecca Strating, Simon Jackman, Max Grömping, Phoebe Hayman, Timothy B Gravelle, '[INFLUENCERS and MESSAGES: Analysing the 2023 Voice to Parliament Referendum Campaign](#)', 17 April 2024.

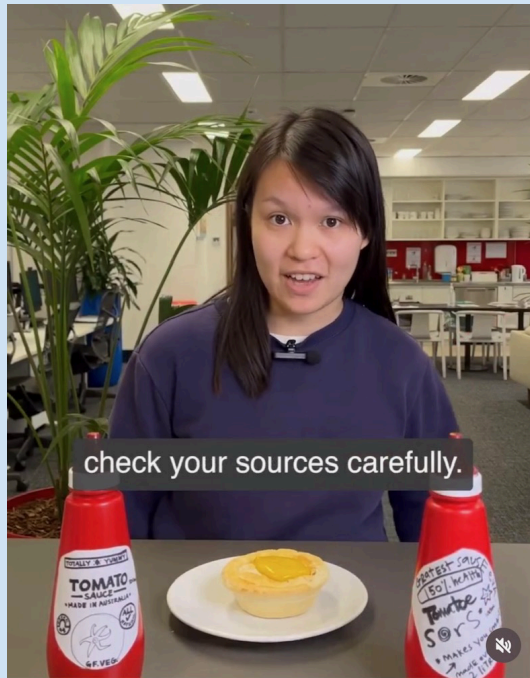


Screenshots of selected videos from the Australian Associated Press media literacy campaign:



77 likes

aapfactcheck If you've ever encountered confusing or contradictory arguments, it may have been a deliberate attempt to mislead you. Incoherence is a common tactic used in disinformation, but being aware of it means you are less likely to fall for it.



40 likes

aapfactcheck Finding sources of information you can trust can be tricky, but good sources will always be open about the information they rely on, their processes, and their mistakes.

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.

- **Meta requires all advertisers of political ads⁴⁵ to complete an ad authorisation, which includes verifying the advertiser's identity.** We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.
- **Meta requires political ads to include a disclaimer disclosing who is paying for the ad.** We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.
- **Meta provides the Ad Library, a searchable archive of all political ads on our services in Australia.** We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

In November 2023, Meta also announced a new AI Disclosure policy to help people understand when a social issue, election or political advertisement on Facebook or Instagram has been digitally created or altered, including through the use of AI.⁴⁶ Advertisers will disclose whenever a social issue, electoral, or political ad contains a photorealistic image or video, or realistic sounding audio, that was digitally created or altered to:

- Depict a real person as saying or doing something they did not say or do; or
- Depict a realistic-looking person that does not exist or a realistic-looking event that did not happen, or alter footage of a real event that happened; or
- Depict a realistic event that allegedly occurred, but that is not a true image, video, or audio recording of the event.

Meta will add information on the ad when an advertiser discloses in the advertising flow that the content is digitally created or altered. This information will also appear in the Ad Library.⁴⁷ If it is determined that an advertiser did not disclose as

⁴⁵ 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".

⁴⁶ Meta, 'Helping People Understand When AI Or Digital Methods Are Used In Political or Social Issue Ads', Meta Newsroom, 8 November 2023.

⁴⁷ Meta, '[Meta Ad library](#)'.

required, Meta will reject the ad. Repeated failure to disclose may result in penalties against the advertiser.

- **Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political ads and social issue ads.**

We have maintained the approach outlined in our 2021, 2022 and 2023 transparency reports.

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.

- **Meta supports research and events in relation to misinformation and media literacy.**

Since the 2023 transparency report, we have supported a number of events and research on misinformation, including:

- In the lead up to the 2023 Australian Aboriginal and Torres Strait Islander Voice Referendum, as mentioned above, we launched a new media literacy campaign in collaboration with the Australia Associated Press building on our “Check The Facts” campaign which ran ahead of the 2022 Federal Election in October 2021 and early 2022. This shared tips and advice with people so that they could make informed decisions about what they read, trust and share.⁴⁸
 - The objective of the campaign was to create engaging, organic-to-the-platform (Facebook and Instagram) content that would give Australian voting age adults the tools to understand what to look out for and what to fact check with the broad aim of limiting the influence misinformation in the period leading up to the referendum. The campaign videos were scripted and shot by AAP FactCheck journalists and aimed to generate awareness of tactics used to spread mis- and disinformation and by driving audiences to check the facts via the AAP Fact Check website.
 - The campaign ran for 6 weeks with a combined reach on Facebook and Instagram in Australia of over 10 million users, creating over 40 million impressions.⁴⁹
- Australian Media Literacy Alliance’s Media Literacy Summit in March 2023.⁵⁰
 - Meta sponsored and hosted the Australian Media Literacy Alliance’s Australian Media Literacy Summit, which brought together a range of journalists, academics, educators, librarians and other experts to discuss and learn about diverse ways to strengthen media literacy education in Australia.

⁴⁸ AAP, ‘[AAP FactCheck and Meta media literacy collab](#)’, AAP Press release, 28 August 2023.

⁴⁹ AAP reporting.

⁵⁰ Australian Media Literacy Alliance, ‘[Australian Media Literacy Summit](#)’, AMLA Event Press Release.

- Meta also supported the Western Sydney University's 2023 Young People and News longitudinal survey
 - This survey provides findings about the news attitudes, practices and experiences of young Australians aged between 8-16 years.
- **Meta will continue to support research and events in relation to disinformation.**

For example, Meta has collaborated with the Australian Strategic Policy Institute (ASPI) for years, sharing our findings and insights about influence operations on our platforms and ingesting leads from networks they've identified.

- **Meta collaborates with researchers to undertake surveys of our users to assess their views on topics such as vaccines and climate change.**

Between March - May 2023 Meta conducted a 'Pandemic Recovery' Survey to inform pandemic response across the area of health education and economic recovery.⁵¹ The survey used aggregated indicators from self reported survey data in more than 21 countries for respondents aged 18 or over.

In 2023 Meta also conducted a 'Climate Change Opinion' Survey, in partnership with Yale, to explore public climate change knowledge, attitudes, policy preferences and behaviours.⁵² This survey included respondents from almost 200 countries and territories.

While Meta will continue to support research that has a positive social impact, our approach to COVID-19 has changed, following feedback from different stakeholders, including Meta's independent Oversight Board. These changes reflect the changing nature and impact of COVID-19 globally since the start of the pandemic. As a result of this, future research topics may vary.

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

The FORT initiative experienced a name change and all relevant transparency efforts can be found on our Transparency Centre.⁵³ Specifically, we have research tools called Meta Content Library and API that share publicly accessible data from Facebook and Instagram. The MCL API is accessible in a secure clean room environment. Meta will make the Meta Content Library and API research tools available to third-party fact-checking partners and qualified users.

⁵¹ Meta, '[Pandemic Recovery Survey](#)', Meta Data for Good Centre.

⁵² Meta, '[Climate Change Opinion Survey](#)', Meta Data for Good Centre.

⁵³ Meta, '[Research Tools and Data Sets](#)', Meta Transparency Centre.

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

- Meta 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.

We met this requirement in 2021, 2022 and 2023 by publishing a transparency report on our Meta Australia blog, and speaking publicly to the media about our work.⁵⁴

The publication of this report satisfies this requirement for 2024.

⁵⁴ Meta, '[Facebook's response to Australia's disinformation and misinformation code](#)', *Meta Australia Blog*, 21 May 2021.

Appendix A: 2023 specific commitments made by Meta under the industry code on disinformation and misinformation

The 38 commitments for 2023 are outlined below:

Outcome 1: Combatting misinformation / disinformation	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. <ul style="list-style-type: none">• Meta removes networks of accounts, Pages and Groups that violate our inauthentic behaviour policy, including disinformation that violates our policy on Inauthentic Behaviour.• Meta provides transparency about accounts, Pages and Groups removed under our Inauthentic Behaviour policy.• Meta partners with experts and organisations who assist in providing tips or further investigation about possible inauthentic behaviour on our services.• Meta removes misinformation that violates the misinformation section of our Community Standards, and will review its Misinformation and Harm Policy, in line with recommendations from the Oversight Board.• Meta removes manipulated videos, also known as “deepfakes”, that violates our Manipulated Media policy.• Meta removes election-related misinformation that may constitute voter fraud or interference under our Coordinating Harm and Promoting Crime policy.• Meta removes fake accounts.• Meta allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.• Meta partners with third-party fact-checking organisations, globally and in Australia, to assess the accuracy of content on our services.• Meta applies a warning label to content found to be false by third-party fact-checking organisations.
--	---

- Meta reduces the distribution of content found to be false by third-party fact-checking organisations.
- Meta proactively searches for content that matches content debunked by our fact-checking partners, to apply the same treatments.
- Meta takes action on Pages, Groups, accounts or websites found to repeatedly share 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.

- Meta makes available a detailed list of claims that we consider to violate our COVID-19 Misinformation & Harm policy.
- Meta makes information available via a dedicated website that outlines our efforts to combat misinformation.

Outcome 1c . Users can report content and behaviours to Signatories that violate their policies under 5.10 through publicly available and accessible reporting tools.

- Meta makes on-platform reporting channels available to users for false information.

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

- Meta makes global transparency reports available regularly.
- Meta will supplement these reports with additional Australia-specific statistics, provided as part of this Annual Report process.

[*NEW*] Outcome 1e. Users will be able to access general information about Signatories' use of recommender systems and have options relating to content suggested by recommender systems.

- Meta will continue to provide greater transparency of our content ranking algorithms and give users more control over the content they see.

	<ul style="list-style-type: none"> • Meta takes steps to limit the possible distribution of misinformation via recommendations.
Outcome 2: Disrupt monetisation and advertising incentives	<ul style="list-style-type: none"> • Meta 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	<ul style="list-style-type: none"> • Meta provides contextual information around posts that users see from public Pages. • Meta provides a Climate Science Information Centre in Australia to connect users to authoritative information from leading climate organisations. • Meta uses in-product prompts to direct Australians to authoritative information on key topics. • Meta promotes authoritative information by providing significant support to organisations such as the Australian, state and territory governments and to promote authoritative health information. • Meta directs users to authoritative information when they search for high-priority topics on Facebook and Instagram. • Meta directs users to authoritative information once they have seen or shared COVID-19 related misinformation. • Meta will look for opportunities to continue to work with the Government on other ways to promote authoritative information. • Meta promotes public service announcements to our users to encourage them to be wary of potential misinformation.

Outcome 5: Political advertising	<ul style="list-style-type: none"> • Meta requires all advertisers of political and social issue ads to complete an ad authorisation, which includes verifying the advertiser’s identity. • Meta requires political and social issue ads to include a disclaimer disclosing who is paying for the ad. • Meta provides the Ad Library, a searchable archive of all political and social issue ads on our services in Australia. • Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political and social issue ads.
Outcome 6: Research	<ul style="list-style-type: none"> • Meta supports research and events in relation to misinformation and media literacy. • Meta supports research and events in relation to disinformation. • Meta collaborates with researchers to undertake surveys of our users to assess their views on topics such as vaccines and climate change. • Meta provides data to researchers in a privacy-protective way via the Facebook Open Research and Transparency initiative.
Outcome 7: Annual reports	<ul style="list-style-type: none"> • Meta 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.