



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

Reporting period: January - December 2025

## Summary

As a founding member and signatory to the Australian Code of Practice on Disinformation and Misinformation (**the Code**), Meta is pleased to share our latest transparency report, highlighting the actions we take to tackle disinformation and misinformation on Facebook and Instagram.

This latest update outlines Meta's approach to combat disinformation and misinformation in Australia during the 2025 calendar year. During the reporting period several trends and issues were identified:

- **Global adversarial threats:** This reporting period, we observed several notable trends by Coordinated Inauthentic Behaviour (CIB) networks, including increased use of Generative AI to rapidly produce convincing text and media at scale; a shift by some foreign actors (including Russian networks targeting Sub-Saharan Africa) toward hiring local, often unwitting freelancers to operate through mature authentic accounts; and the increasingly cross-internet nature of campaigns coordinating across multiple platforms and off-platform sites. Despite these adaptations, our detection systems remained robust in their ability to detect CIB activity and most disrupted networks continued to struggle to build genuine audiences. We also saw persistent tactical evolution by recidivist operations such as Russia's Doppelganger (including a pivot toward posting largely unreadable content in fruitless attempts to evade our enforcement efforts) and Iran's Endless Mayfly (including movement from typosquatting<sup>1</sup> to bespoke outlets and frequent infrastructure migration). Finally, in response to an emerging threat vector involving potential data-poisoning<sup>2</sup> efforts - where influence operations design content and websites to shape AI training data toward specific narratives—we blocked policy

---

<sup>1</sup> Typosquatting: a type of social engineering attack which tricks users into visiting malicious websites with URLs that are common misspellings of legitimate websites. Meta, '[Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2025](#)', Meta Transparency Centre, December 2025

<sup>2</sup> Data poisoning: Manipulating training data to subvert model behavior and/or introduce vulnerabilities. Meta, '[Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2025](#)', Meta Transparency Centre, December 2025

violating sites from our training data and retrieval-augmented generation (RAG) processes.<sup>3 4 5</sup>

- **Continued further deployment of Generative AI:** We began gradually rolling out “AI Info” labels for certain ad creative images and videos using a risk-based framework. When an image or video is created or significantly edited with our generative AI creative features in our advertiser marketing tools, a label will appear in the three-dot menu or next to the “Sponsored” label. We plan to share more information on our approach to labelling ad images made or edited with non-Meta generative AI tools. We will continue to evolve our approach to labelling AI-generated content in partnership with experts, advertisers, policy stakeholders and industry partners as people’s expectations and the technology change<sup>6</sup>.
- **Rapid evolution of adversarial tactics for fraud and scams:** This reporting period, we continued to see bad actors rapidly evolve and adapt to evade detection. We outlined our multi-layered approach to scam reduction, including using automated technical defenses and AI-driven detection, disrupting scam networks, expanding advertiser verification efforts (with particular focus on higher-risk areas such as financial investment ads), and building product features to help people spot and avoid suspicious engagement. We also underscored the importance of collaboration beyond our platforms, including working with cross-industry partners and law enforcement to share signals, support investigations, and take action against transnational criminal enterprises behind many scam operations.<sup>7</sup>

This is Meta’s sixth transparency report released under the Code since it took effect in 2021. Our last report outlined 39 commitments to meet our obligations across both Facebook and Instagram for the reporting period of January to December 2025.<sup>8</sup> This report outlines the steps we took to meet these commitments.

For this reporting period, we are releasing up-to-date, Australia-specific data on our misinformation efforts. *Note: We expanded the scope of our “Misinformation” policy category to also include coordinating harm and promotion of crime, and inauthentic behavior policy violations to align with our EU report for the Code of Practice on Disinformation. As a result, the total actions enforced for content under this category increased materially and are not directly comparable to 2024 figures.*

Between 1 January and 31 December 2025:

- We took action on over 270,000 pieces of content across Facebook and Instagram in Australia for violating our Misinformation policies.
- In addition to this, over 9,700 ads were removed in Australia for violating our Advertising Standards on Misinformation.
- We displayed warnings on over 4.8 million distinct pieces of content on Facebook, and over 236,000 on Instagram in Australia (including reshares), based on articles written by our third-party fact-checking partners.

---

<sup>3</sup> Meta, [‘Meta’s Adversarial Threat Report, First Quarter 2025’](#), Meta Transparency Centre, May 2025

<sup>4</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

<sup>5</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, First Half 2026’](#), Meta Transparency Centre, March 2026

<sup>6</sup> Meta, [‘Expanding GenAI Transparency for Meta’s Ads Products’](#), February 2025

<sup>7</sup> Meta, [‘Scams Are Bad for Business: Our Ongoing Efforts to Fight Fraud’](#), December 2025

<sup>8</sup> This reporting period has been specified in the guidelines provided by the independent reviewer, engaged by DIGI

- We removed over 65,000 ads in Australia for not complying with our social issues, elections and politics (SIEP) ads policy.

In 2025, some further examples of our work in Australia included:

In the lead-up to, during and after the 2025 Australian Federal Election, Meta implemented a multi-layered approach to safeguard information integrity across Facebook and Instagram. Further details about our work around the 2025 election is detailed later on in this report.

Meta supported research led by La Trobe University and the University of Waikato that analysed over three million posts from 25 major Australian news publishers spanning 15 years. The study revealed that arts and lifestyle content now leads audience engagement on Facebook, reversing a decade-old pattern of political content dominance (see Commitment 35).<sup>9</sup>

Partnering with Australian Associated Press (AAP) we delivered a multi-literacy campaign (in English, Simplified Chinese, Vietnamese and Arabic) that reached 823,638 people in Australia, accumulated over 5.07 million views, and generated more than 2.45 million completed video plays.<sup>10</sup>

---

<sup>9</sup> Andrea Carson, ['Facebook and the News Media: How Australians Engage with News and Misinformation Online'](#), September 2025

<sup>10</sup> Campaign results supplied by Australian Associated Press.



In April 2025, we hosted a hybrid roundtable event at the Meta office in Sydney, titled 'Supporting Media Literacy and Information Integrity in the AI Age'. The event convened a group of government officials, regulators, and academic, industry and other experts focused on information integrity for an open discussion as Australia was preparing for the May 2025 federal election and as the use of generative AI technology was becoming increasingly prevalent. The discussion explored strategies to mitigate the potential risks for AI technology to be misused to spread or amplify mis- and disinformation; reflected on some of the lessons from the 2024 global elections; and addressed the importance of transparency and media literacy education as AI technology becomes increasingly sophisticated.

In September 2025, Meta partnered with DIGI to co-host the annual Australian Code of Practice on Disinformation and Misinformation event, titled 'Information Integrity in the Gen AI Era'. The event brought together over 50 representatives from government, academia, civil society and

industry to explore the information integrity challenges and opportunities presented by the growing use of generative AI technology in everyday life.

In the year ahead, 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 furthering our collaboration with Australian policymakers, civil society, academics and experts to take meaningful steps against misinformation and disinformation in Australia.

## Table of contents

Summary	1
Table of contents	6
Outcome 1a	7
Case study: Meta’s response to global CIB attempts and global adversarial threats throughout 2025	8
Case study: Meta’s 2025 Australian Election Integrity Response	12
Outcome 1b	17
Outcome 1c	18
Outcome 1d	19
Outcome 1e	20
Outcome 2	22
Outcome 3	22
Outcome 4	23
Outcome 5	25
Outcome 6	28
Outcome 7	30
<b>Appendix A: 2025 specific commitments made by Meta under the industry code on disinformation and misinformation</b>	<b>31</b>

## Reporting against 2025 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**

During the reporting period, our approach to misinformation and disinformation in Australia is consistent with that which was outlined in our previous transparency reports.<sup>11</sup> Below we outline developments from 2025:

- **Commitment 1: Meta removes networks of accounts, Pages and groups that violate our Inauthentic Behaviour (IB) policy, including disinformation networks that violate our policy on Coordinated Inauthentic Behaviour (CIB).**

We have also maintained the approach to misinformation outlined in our previous transparency reports. Meta's approach to misinformation focuses on removing content which could directly contribute to the risk of imminent physical harm, interfere with the functioning of political processes, or constitute highly deceptive manipulated media, as well as removing fake accounts which are often vehicles for disinformation. It also focuses on reducing the spread of other misinformation, while promoting authoritative information.

- **Commitment 2: 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 IB and CIB - both foreign and domestic.

We continue to report on our efforts to disrupt fake accounts and CIB through our Community Standards Enforcement Report and Adversarial Threat Reports.<sup>12</sup>

Meta removed 16 new covert influence operations globally in 2025.<sup>13 14 15</sup> The majority of the CIB networks we've disrupted struggled to build genuine followings. They typically attempted to spread their activities across multiple online platforms, including our own, as well as YouTube, TikTok, and others. Notably, some have shifted their focus to

---

<sup>11</sup> Meta, ['Meta's response to the Australian disinformation and misinformation industry code'](#), Digi, May 2025

<sup>12</sup> Meta, ['Community Standards Enforcement Report'](#), and Meta, ['Adversarial Threats reports'](#), Meta Transparency Centre, May 2025

<sup>13</sup> Meta, ['Meta's Adversarial Threat Report, First Quarter 2025'](#), Meta Transparency Centre, May 2025

<sup>14</sup> Meta, ['Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2025'](#), Meta Transparency Centre, December 2025

<sup>15</sup> Meta, ['Meta's Semiannual Adversarial Threat Report, First Half 2026'](#), Meta Transparency Centre, March 2026

platforms with fewer safeguards, given continuous enforcement efforts on our services.

Below we provide a case study that looks at the adversarial threat trends globally across 2025.

## Case study: Meta's response to global CIB attempts and global adversarial threats throughout 2025

A high level summary of CIB Networks disrupted in 2025:

- In 2025, 16 new covert influence operations were removed.
- In the First Quarter (Q1 2025) we disrupted three separate covert influence operations for violating our policy against CIB. They originated in:
  - China (targeting Myanmar, Taiwan, and Japan),
  - Iran (targeting Azeri-speaking audiences in Azerbaijan and Turkey), and
  - Romania (targeting domestic audiences).
- In the Second and Third Quarters (Q2 - Q3 2025) we disrupted seven separate covert influence operations that violated our policy against CIB. They originated in:
  - Belarus (targeting Poland)
  - India (domestic)
  - Moldova (domestic), and
  - Russia (targeting Moldova and Sub-Saharan Africa),
- In the fourth quarter (Q4 2025) we removed six separate covert influence operations that violated our policy against CIB. They originated in:
  - China (targeting Taiwan),
  - Iran (targeting Azerbaijan as well as US and English-speaking audiences),
  - Pakistan (domestic), and
  - Russia (targeting Sub-Saharan Africa and approximately 20 countries).

### Trends in coordinated inauthentic behaviour throughout 2025:

Throughout 2025, coordinated inauthentic behavior (CIB) networks evolved their tactics to evade detection and attempt to distribute content on our services. Several major trends defined the CIB landscape in 2025:

- **Proliferation and Integration of Generative AI:** Generative AI facilitated scaling for threat actors, allowing them to generate content without needing specialized writers or translators.<sup>16</sup> Across various networks originating from China, India, Russia, and Pakistan, operators used AI to:

<sup>16</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2026](#)', Meta Transparency Centre, December 2025

- Generate highly convincing profile pictures and biographies for fake personas.<sup>17 18</sup>
  - Craft polished, multi-lingual text for posts, comments, and tailored news articles.<sup>19 20</sup>
  - Create sophisticated multimedia content, including photorealistic AI-generated video of “journalist” personas conducting fake interviews, as seen in a domestic influence operation in Pakistan.<sup>21</sup>
- It’s important to note that despite these benefits for scaling, we have yet to see an increase in threat actors’ ability to evade our detection. This is because our enforcements against persistent threat actors running CIB operations have long relied on behavioral and technical signals, rather than content signals.
  - **Two-Tiered Structures and Deep Persona Development:** Instead of relying solely on mass bot networks, threat actors invested in creating “custom personas” with deep backstories, including specific professions, education histories, and culturally symbolic imagery.<sup>22 23</sup>
  - **Outsourcing to Potentially Unwitting Local Freelancers** In a notable tactical shift, influence operations increasingly moved away from using exclusively inauthentic accounts, opting instead to hire potentially unwitting local freelance social media managers.<sup>24</sup> This was particularly evident in Russia-linked campaigns targeting Sub-Saharan Africa.<sup>25</sup> By outsourcing content distribution to authentic, locally based digital media managers, foreign actors attempted to minimize their online footprint, obtain mature local infrastructure, and smoothly integrate their campaigns into the local information environment.<sup>26</sup>
  - **Grassroots Media Impersonation and Co-opting Authentic Content** CIB networks consistently disguised themselves as local, grassroots media outlets or

<sup>17</sup> Meta, [‘Meta’s Adversarial Threat Report, First Quarter 2026’](#), Meta Transparency Centre, May 2025

<sup>18</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

<sup>19</sup>Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

<sup>20</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, First Half 2026’](#), Meta Transparency Centre, March 2026

<sup>21</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, First Half 2026’](#), Meta Transparency Centre, March 2026

<sup>22</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2026’](#), Meta Transparency Centre, December 2025

<sup>23</sup>Meta, [‘Meta’s Semiannual Adversarial Threat Report, First Half 2026’](#), Meta Transparency Centre, March 2026

<sup>24</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

<sup>25</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

<sup>26</sup> Meta, [‘Meta’s Semiannual Adversarial Threat Report, Second and Third Quarter 2025’](#), Meta Transparency Centre, December 2025

activist groups.<sup>27 28</sup> Operations used "typosquatting" to mimic legitimate news websites, set up standalone off-platform blogs and impersonated journalists.<sup>29 30</sup>

Additionally, instead of relying purely on fabricated content, networks increasingly attempted to co-opt authentic voices. For example, an Iranian network reshared posts from real, high-profile influencers while adding their own captions to appear legitimate, while another network injected their narratives directly into the comment sections of prominent authentic news pages like the BBC and RadioLiberty.<sup>31</sup>

- **Commitment 3: 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.

- **Commitment 4: Meta removes misinformation that violates the misinformation section of our Community Standards.**
- **Commitment 5: Meta removes misinformation that violates our policy against posting content that directly contributes to a risk of interference with participation in election processes.**
- **Commitment 6: Meta removes content that violates our policies prohibiting Coordination of Harm and Promoting Crime that calls for violence, coordinated interference in elections or promotes crime.**

For the reporting period, in Australia we have maintained the approach outlined in our previous transparency reports. Between 1 January and 31 December 2025, we took action on over 270,000 pieces of content across Facebook and Instagram in Australia for violating our Misinformation policies.<sup>32</sup>

We continue to remove content that may cause voter suppression or misrepresentation of election-related information. Specifically, our policies prohibit misrepresentation of the date, locations, times or methods of voting or voter registration (such as claims that voting can be done via an online app); misrepresentation of who can vote, qualifications for voting, whether a vote will be counted, or what information or materials must be provided in order to vote; and statements that advocate, provide instructions for, or show explicit intent to illegally participate in a voting process. We recognise that voting is

---

<sup>27</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, First Half 2026](#)', Meta Transparency Centre, March 2026

<sup>28</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2025](#)', Meta Transparency Centre, December 2025.

<sup>29</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, Second and Third Quarter 2025](#)', Meta Transparency Centre, December 2025

<sup>30</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, First Half 2026](#)', Meta Transparency Centre, March 2026

<sup>31</sup> Meta, '[Meta's Semiannual Adversarial Threat Report, First Half 2026](#)', Meta Transparency Centre, March 2026

<sup>32</sup> Note: This figure reflects an expanded policy scope for the 2025 reporting period (see Summary for details) and is not directly comparable to prior years.

essential to democracy, and accordingly take a stricter approach to misrepresentations and misinformation that could result in voter fraud or interference.

In 2024, we introduced political content controls on Facebook and Instagram to give users more choice over how much political content is recommended to them. As of May 2025, these controls have been rolled out globally.<sup>33</sup>

We also updated our penalty protocol related to public figures suspended for violations during periods of civil unrest. Periodic reviews are held to assess whether enhanced penalties are warranted.

Meta remains focused on providing users with reliable election information while combatting misinformation across languages. Meta continues to connect users with details about elections through in-app notifications on Facebook and Instagram. These reminders included information on registering to vote, voting by mail, voting early in person, and voting on election day.

### **Case study: Meta’s 2025 Australian Election Integrity Response**

- Ahead of Australia’s May 2025 Federal Election, Meta continued our longstanding partnership with the Australian Electoral Commission across a number of areas. This included activating our voter empowerment products, which helped to connect Australians with verified election information and voting reminders. For those who wished to share their civic experience, Instagram voting stickers were also available for people to post to their Stories.
- We also activated a cross-functional team to help ensure the integrity of the election on our platform - one that gave people a voice, supported participation in the civic process, and worked to combat voter interference and foreign influence on our services.
- To combat misinformation, we partnered with Agence France-Presse (AFP) and Australian Associated Press (AAP) for independent fact-checking, applying warning labels and reducing distribution of debunked content so that it was less likely to be seen.
- We also launched a media literacy campaign with AAP (see Commitment 35 for full details) which reached over 823,000 Australians and generated more than 5 million views across English, Simplified Chinese, Vietnamese, and Arabic formats.
- We continued to provide transparency for social issues, elections or politics ads, according to our advertising policies. Advertisers who run these ads are required to complete an authorisation process and include a “paid for by”

<sup>33</sup> Meta, [‘Our Approach to Political Content | Transparency Center’](#), September 2025

disclaimer. These ads are then stored in our publicly available Ad Library for seven years.

- **Commitment 7: 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.

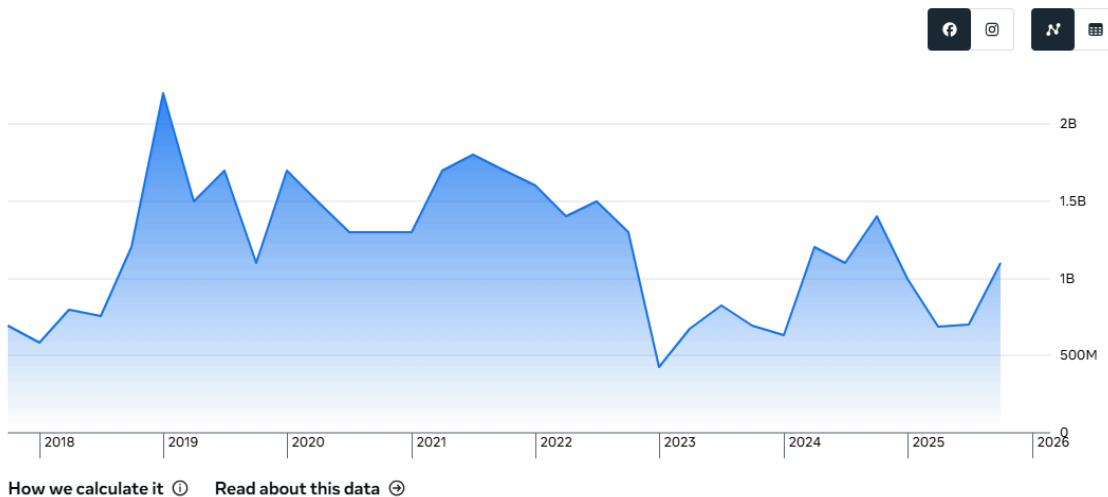
Our ability to detect and remove fake accounts has been improving over the years, and there has been a general decline in the volume of fake accounts found on the platform since 2019.

From January to December 2025, we detected and removed 3.4 billion fake accounts on Facebook, on average we proactively detected and removed over 99% of these accounts before they were reported to us.<sup>34</sup>

### No. of fake accounts we've taken action on (Oct 2018 - Dec 2025)<sup>35</sup>

ACCOUNTS ACTIONED

How many fake accounts did we take action on?



- **Commitment 8: Meta allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.**

As mentioned in our transparency reports from 2021 to 2025, Meta has voluntarily established the independent Oversight Board to ensure greater accountability for our

<sup>34</sup> Meta, '[Community Standards Enforcement Report](#)', Meta Transparency Centre, May 2026

<sup>35</sup> Note - graph represents number of fake accounts removed globally. Australia-specific statistics are not available.

content governance on Facebook and Instagram.

Since its first decision in 2021, the Board has issued 335 recommendations to Meta. By the end of 2025, Meta had publicly responded to 326 of these recommendations, committing to implement or explore the feasibility of 80% of them.<sup>36</sup>

Examples of the Oversight Board's recommendations and guidance areas relevant to misinformation include:

- **Election Integrity Pillars:** While stemming from an earlier Oversight Board recommendation concerning Brazil's election, Meta operationalised this guidance in 2025 by formally establishing eight core election pillars.<sup>37</sup> Two of these pillars - "Preventing interference and disinformation" and "Reducing the spread of misinformation" - are explicitly dedicated to mitigating misinformation.
- **Consistent Labelling for Manipulated Media:** In the second half of 2025, the Oversight Board published a significant decision regarding the audio clip that allegedly captured Kurdish politicians discussing plans to rig parliamentary elections in Iraqi Kurdistan. Meta had initially applied "digitally altered" labels to some instances of this audio but left the specific case content unlabelled to avoid false positives. The Oversight Board overturned our decision, emphasising that information labels must be applied consistently to potentially manipulated media, particularly because information integrity is paramount during elections. In response to this, we updated our internal guidance and improved our practices to ensure prominent labels are consistently applied to manipulated media, including duplicate content.<sup>38</sup>
- **Commitment 9: Meta has partnered with third-party fact-checking organisations in Australia, to assess the accuracy of content on our services.**

As noted above, this report addresses the practices implemented for Facebook and Instagram in Australia during the reporting period of January to December 2025. In alignment with Meta's public announcements on 7 January 2025,<sup>39</sup> we continue to evaluate the applicability of these practices to Meta products. We also regularly review the appropriateness of making adjustments in response to changes in our practices, such as the deployment of Community Notes. To date, Community Notes have launched in the US only, with no plans to launch in Australia at the time of this report's publication.

During the reporting period of January to December 2025, under the third-party fact-checking program, Meta continued to partner with two third-party fact-checkers in relation to Australia content: Australian Associated Press (AAP) and Agence France Presse (AFP).

---

<sup>36</sup> Meta, ['Meta's Bi-Annual Reports on the Oversight Board'](#), March 2026

<sup>37</sup> Meta, ['Meta's Bi-Annual Reports on the Oversight Board'](#), August 2025

<sup>38</sup> Meta, ['Meta's Bi-Annual Reports on the Oversight Board'](#), March 2026

<sup>39</sup> Meta, ['More Speech and Fewer Mistakes'](#), Meta Newsroom, January 2025

- **Commitment 10: Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta applies a warning label to content found to be false, partly false, altered, or missing context by third-party fact-checking organisations.**

In relation to the reporting period, in Australia we have maintained the approach outlined in our previous transparency reports. 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 and Instagram. 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.<sup>40</sup>

- **Commitment 11: Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta reduces the distribution of content found to be false, partly false, or altered, by third-party fact-checking organisations.**

Meta has maintained the approach outlined in our previous transparency reports. Where content is rated as false, we reduce its distribution so that it appears lower in Facebook Feed and is less discoverable on Instagram - including removal from Explore and hashtag pages, and downranking in Feed and Stories. This significantly slows the spread of misinformation.

- **Commitment 12: Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta proactively searches for content that matches content debunked by our fact-checking partners, to apply the same treatments.**

During the reporting period of January to December 2025, Meta used our technology to detect content that was the same or near-identical versions to content that was 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 and December 2025, we displayed warnings in Australia on over 4.8 million distinct pieces of content on Facebook, and over 236,000 on Instagram, in Australia (including re-shares) based on articles written by our third-party fact-checking partners.

- **Commitment 13: Meta takes action on Pages, groups, accounts, or websites found to repeatedly share misinformation.**

---

<sup>40</sup> Meta, ['How fact-checking works'](#), May 2026

Meta has maintained the approach outlined in our previous transparency reports. 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 monetize and advertise and removing their ability to register as a news Page.<sup>41</sup>

- **Commitment 14: Meta safeguards into our GenAI features and models so that people can have a safer and more enjoyable experience.**

Meta has continued to work across many channels to combat deceptive use of AI in 2025 to align with its goals, including through public reporting, contributions to industry coalitions and standards organisations, releasing of open source tools, and a variety of mechanisms in the user-facing surfaces of our products and services. We are continuing to follow through on our Accord commitments in 2025.<sup>42</sup>

Since the start of 2024, in certain cases, advertisers must also now disclose when they use generative AI to create or alter a political or social-issue ad. This disclosure obligation applies if the 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.

It also applies if an ad depicts a realistic-looking person that does not exist or a realistic-looking event that did not happen, alters footage of a real event, or depicts a realistic event that allegedly occurred, but that is not a true image, video, or audio recording of the event. We also add information on the ad when an advertiser discloses that the content has been digitally created or altered. This information also appears in the Meta Ad Library.

When Meta determines that an advertiser running SIEP (social issues, elections and politics) ads has not disclosed digitally created or altered content where required, Meta rejects the relevant ad. Repeated failure by an advertiser to disclose may result in penalties against the advertiser.

We believe it is important to inform our users where content is AI-generated, so all content created with Meta AI is labelled as “AI Info”. We also add “AI Info” labels to content when we detect industry standard metadata or when people disclose that they are uploading AI-generated images, video, or audio. We require people to disclose, using our AI-disclosure tool, whenever they post organic content with photorealistic video or realistic-sounding audio that was digitally created or altered, and we may apply penalties if they fail to do so.

Our Community Standards apply to all content on our services, regardless of whether it was created by AI technology.

---

<sup>41</sup> Meta, [‘How fact-checking works’](#), May 2026

<sup>42</sup> Meta’s progress reporting under the Tech Accord to Combat Deceptive Use of AI is available via the AI Elections Accord [website](#).

- **Commitment 15: Meta provides tools to help people know when they see content that was made with AI.**

In addition to the measures outlined in response to Commitment 14, AI-generated content is eligible to be reviewed and rated by Meta's third-party fact-checking partners, whose rating options<sup>43</sup> allow them to address various ways in which media content may mislead people, including but not limited to media that is created or edited by AI.

Meta labels content when we detect industry-standard AI image indicators or when users disclose that they're uploading AI-generated content. Meta requires people to use this disclosure and label tool when they post organic content with a photorealistic video or realistic-sounding audio that was digitally created or altered, and may apply penalties if they fail to do so. If Meta determines that digitally created or altered image, video, or audio content creates a particularly high risk of materially deceiving the public on a matter of importance, we may add a more prominent label, so that people have more information and context.

### **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.**

- **Commitment 16: Meta makes information available via a dedicated website that outlines our efforts to combat misinformation.**

We have maintained the approach outlined in our previous transparency reports. Meta regularly publishes updates about our misinformation policies and procedures via our Newsroom (<https://about.meta.com/news>). These updates have covered initiatives such as connecting people with authoritative information, updating advertising policies, removing false claims, launching dedicated information hubs, supporting fact-checkers and local news organisations, and limiting the spread of harmful misinformation across our apps. Meta's policies on misinformation can be found in our Transparency Centre.<sup>44</sup>

### **Outcome 1c**

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

<sup>43</sup> Meta, '[Rating options for fact-checkers](#)', May 2026

<sup>44</sup> Meta, '[Community Standards - Misinformation](#)', Meta Transparency Centre. May 2026

**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.**

- **Commitment 17: 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. On both platforms, users are able to report content directly in-app by selecting the menu on any post and choosing a reporting category for “sharing false information”. On Facebook reporting is also available for Pages and groups. We recommend on-platform reporting as the primary channel, as it allows us to identify the content in question and ensures a simpler, more timely process for users without requiring them to leave their user experience. Steps to report misinformation can be found at the following Help Centre links:

- Facebook Help Centre - <https://www.facebook.com/help/572838089565953>
  - Instagram Help Centre - <https://www.facebook.com/help/instagram/2442045389198631>.
- **Commitment 18: Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta's Australian third-party fact-checking partners are also able to receive referrals from the public.**

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/factcheck/>

## **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.**

- **Commitment 19: 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.<sup>45</sup>

<sup>45</sup> Meta, '[Meta Transparency Reports](#)', Meta Transparency Centre, May 2026

Meta publishes quarterly metrics on the detection and enforcement of content that violates our Community Standards. These metrics are available in Meta’s Community Standards Enforcement Report.<sup>46</sup>

In 2022, Meta published the results of an independent audit, conducted by Ernst & Young, into the Community Standards Enforcement Report. Ernst & Young 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.<sup>47</sup> We continued publishing our Community Standards Enforcement Report throughout 2025.

- **Commitment 20: Where available and useful, 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, for the reporting period Meta is providing Australia-specific statistics.

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 Advertising Standards on Misinformation.

These statistics can be found in the relevant commitment sections of this report: content enforcement actions (Commitments 4-6, under Outcome 1a), fact-checking warning labels (Commitment 12, under Outcome 1a), and advertising enforcement (Commitment 23, under Outcome 2).

## **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.**

---

<sup>46</sup> Meta, [‘Community Standards Enforcement Report’](#), Meta Transparency Centre, May 2026

<sup>47</sup> Meta, [‘Community Standards Enforcement Report’](#), Meta Transparency Centre, May 2026. The results of the EY assessment can be found at [REPORT OF MANAGEMENT ON THE INTERNAL CONTROLS OVER THE CALCULATION AND REPORTING OF THE FACEBOOK AND INSTAGRAM COMMUNITY STANDARDS](#), May 2022

**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.**

- **Commitment 21: Meta will continue to provide ongoing transparency of our content ranking algorithms and give users more control over the content they see.**

In December 2025, Meta launched in the U.S. "Your algorithm" on Instagram — an AI-powered feature that gives users direct, meaningful control over their Reels recommendations. In January 2026, it was rolled out to English-speaking markets (excluding South Korea and India). The feature allows users to actively view, manage and personalise the content they see in Reels.

"Your algorithm," also known as tune your algorithm (TYA), is an AI-powered feature that provides users with the ability to control and personalise the content they see in their Reels recommendations experience. This feature is part of the broader AI Interests system and represents a meaningful new way to shape what users see, particularly as recommendations continue to grow.

TYA functions through three main components: Inference, which converts engagement into inferred interests; Control, where users can view and edit these interests (adding topics they want to see "more" of or "less" of); and Reflection, where these interests influence the content served across recommendation surfaces.

TYA allows users to actively manage their algorithm by adding and removing specific interests. Users enter the feature through the top right in the Reels tab and see a summary ("Lately" string) along with editable, automatically inferred interests.

The system is engineered to be safe, leveraging IPGen (a framework for AI-generated content) to enforce Community Standards, Non-Rec policies, Content Risk Standards (CRS), and Sensitive topics at the interest inference and generation stage.

This feature works in addition to content preferences a user has shared with us. People can already express they're "Interested" or "Not interested" in content they see on feed by clicking the three-dot menu. With advancements in AI, we are now able to give people even more control over what they see on Instagram.

Our policies do not allow topics that violate our Community Standards to be added to "What you want to see more of." If a user attempts to input a problematic term, they will see a pop-up explaining that the topic cannot be added.

The interests a user adds in Your algorithm may influence the ads they see, as ads can be based on interaction with content recommended in the Reels feed. However, the feature does not allow users to directly request to see more or fewer ads.

- **Commitment 22: Meta takes steps to limit the possible distribution of misinformation via recommendations.**

Meta takes steps to limit the possible distribution of misinformation via recommendations. From January to December 2025, Facebook and Instagram content that was debunked<sup>48</sup> by non-partisan, third-party fact-checking organisations who partner with Meta, and certified by the International Fact-Checking Network (IFCN) was removed from recommendations.<sup>49</sup>

Pages, groups, Profiles, websites, and Instagram accounts that repeatedly shared content rated False or Altered were put under some restrictions for a given time period. This included removing them from the recommendations we show people, reducing their distribution, removing their ability to monetise and advertise.<sup>50</sup>

## 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.**

- **Commitment 23: 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 previous transparency reports. Meta requires all advertisers to comply with our Advertising Standards, which go beyond our Community Standards for organic content. We prohibit ads that include content debunked by third-party fact-checkers, and advertisers that repeatedly post information deemed to be false may have restrictions placed on their ability to advertise across Meta technologies.

We removed over 9,700 ads for violating our misinformation policy from January to December 2025 in Australia. We also removed over 65,000 ads for not complying with SIEP ads policies from January to December 2025 in Australia.

- **Commitment 24: Meta commits to prevent fact-checked content from being advertised or monetised**

Meta uses a combination of enforcement technology, human review and independent fact-checkers to identify, review and take action on this type of content.

<sup>48</sup> Meta, [‘About fact-checking on Facebook, Instagram and Threads’](#), Meta Business Help Centre, May 2026

<sup>49</sup> Meta, [‘Fact-checked Misinformation’](#), Meta Transparency Centre, May 2026

<sup>50</sup> Meta, [‘Penalties for sharing fact-checked content’](#), Meta Transparency Centre, May 2026

Ads must comply with the Advertising Standards. Meta prohibits ads that include content debunked by third-party fact-checkers. Advertisers that repeatedly post information deemed to be false may have restrictions placed on their ability to advertise across Meta technologies<sup>51</sup>.

### **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 Coordinated Inauthentic Behavior.

### **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.**

- **Commitment 25: 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 previous transparency reports in terms of the contextual information around posts that users see from public Pages.

Meta provides a range of labels, signals, and contextual tools to help users assess the trustworthiness of content they encounter. These included a context button that provides information about the sources of articles in News Feed, notifications alerting users when articles they are about to share are more than 90 days old, and labels identifying content from news organisations that are partially or fully under the control of their government (available on both Facebook and Instagram). We also provide transparency about Pages through the Page Transparency tool which displays information such as previous Page names, the number of Page admins, the countries in which admins are located, and ads currently being run by the Page.

---

<sup>51</sup> Meta, [‘Advertising Standards’](#), Unacceptable Content, Misinformation, May 2026

- **Commitment 26: 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.**

An effective way to address online safety and harmful content is to build a resilient digital society by providing the tools and resources that will enable people to make informed decisions. By empowering individuals with the skills to critically evaluate information, we can foster a culture where people have a better ability to be effective in deciding what to read, trust, and share online. We do this by providing greater transparency and control to users; providing information that will help them make informed decisions; and advancing media and digital literacy.

We offer many tools, products and resources to users to address different areas of safety risks and harms, including:

- Authoritative information sources
  - Safety hubs
  - Warning labels and notices
  - Parental supervision and age-appropriate controls
  - Comments filtering tools
  - Context buttons with more information
  - Privacy tools
  - Controls to customise what users see in their Feed
  - Feed options that allow users to decide how they want content ranked.
- **Commitment 27: Meta reduces the distribution and recommendation of problematic and low quality content on Facebook and Instagram.**

Meta’s Content Distribution Guidelines describe the types of content we think may either be problematic or low quality, as such, we reduce its distribution in Feed for everyone.<sup>52</sup>

- **Commitment 28: Meta provides transparency and information for users in relation to AI-generated content, including through developing industry standard indicators for labelling AI-generated content**

Please see Outcome 1a for the actions Meta takes in relation to AI-generated images, videos, and audio content.

- **Commitment 29: Meta will look for opportunities to continue to work with the Government on other ways to promote authoritative information.**

Meta continues to collaborate with Australian government and institutional partners to promote authoritative information.

In 2025, ahead of the Australian Federal Election, we partnered with the Australian Electoral Commission to help ensure the integrity of election-related information on Facebook and Instagram before, during and after the election. Meta deployed a number of products across Facebook and Instagram designed to connect Australians with

---

<sup>52</sup> Meta, [‘Types of Content We Demote’](#), Meta Transparency Center, July 2025

authoritative voting information from the Australian Electoral Commission and encourage civic participation. These products included Voter Information Units on Facebook and Instagram, which ran in the week prior to election day to direct users to authoritative information about how, when and where to vote. On election day itself, we deployed Election Day Reminders on Facebook and Instagram to encourage people to vote.

These products reached a significant number of Australians:

- Our Voter Information Unit reached around 11.8 million users on Facebook and around 9.5 million users on Instagram.
- Our Election Day Reminder reached around 9.9 million users on Facebook and around 7.5 million users on Instagram.

Meta also partnered with AAP to deliver a multilingual media literacy campaign that reached over 823,000 Australians ahead of the federal election (see Commitment 35 for further details).

## 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.**

During the reporting period, Meta removed over 65,000 ads in Australia for not complying with our social issues, elections and politics (SIEP) ads policy. This enforcement activity underpins the authorisation, disclosure, and transparency measures described below.

- **Commitment 30: Meta requires all advertisers of political and social ads<sup>53</sup> <sup>54</sup> to complete an ad authorisation, which includes verifying the advertiser's identity.**

We have maintained the approach outlined in our previous transparency reports. Meta requires any advertiser in Australia running political or social issue ads to complete a prior authorisation process, which includes verification of the advertiser's identity using official identity documentation. We periodically check for accuracy, and if provided information appears to be invalid or unavailable, the relevant ads will be taken down.

- **Commitment 31: Meta requires political and social issue ads to include a disclaimer disclosing who is paying for the ad.**

We have maintained the approach outlined in our previous transparency reports. To help journalists, civil society, and the general community understand who is behind political and social issue ads, Meta requires these ads to include a disclaimer that accurately reflects the organisation or person paying for them.

<sup>53</sup> We define ads about social issues, elections and politics as advertisements: (1) made by, on behalf of or about a candidate for public office, a political figure, a political party, a political action committee 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; or (3) about social issues in any place where the ad is being placed; or (4) regulated as political advertising".

<sup>54</sup> Meta, ['About ads about social issues, elections or politics'](#), Business Help Centre, May 2026

- **Commitment 32: Meta provides the Ad Library, a searchable archive of all political and social issue ads on our services in Australia.**

We have maintained the approach outlined in our previous transparency reports.

Advertisers running political and social issue ads must complete an authorisation process and include a "paid for by" disclaimer. These ads are stored in our publicly accessible Ad Library for seven years.

Advertisers are also required to disclose when they use AI techniques to create or materially alter a political or social issue ad.

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.<sup>55</sup> If it is determined that an advertiser did not disclose as required, Meta will reject the ad. Repeated failure to disclose may result in penalties against the advertiser.

- **Commitment 33: Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political and social issue ads.**

We have maintained the approach outlined in our previous transparency reports. Meta makes available the Ad Library Report which allows for aggregated analysis of political and social issue ads on Facebook and Instagram, allowing for easy comparison and analysis between advertisers.

- **Commitment 34: Meta provides transparency and information for users in relation to AI-generated content used in ads.**

Meta continues to expand our transparency measures for AI-generated content in advertising, building on our commitment to providing users with clear information about how AI is used in advertisements.<sup>56</sup> Since 2024, we have been applying a visible "AI Info" label to image and video ads created or significantly edited using our in-house AI ad creative features, and in 2025 we further strengthened and broadened these measures. We remain committed to transparency and responsibility in our advertiser guardrails, and continue to collaborate with experts, policy stakeholders, and industry partners to shape our evolving approach.

At the core of our transparency framework is the visible "AI Info" label, which is applied to ad images and videos that have been created or significantly edited using our in-house ad creative AI tools. In 2025, Meta expanded this labelling across more surfaces. When an image or video is created or significantly edited with our generative AI creative features in our advertiser marketing tools, a label will appear in the three-dot menu or next to the "Sponsored" label, ensuring users are informed about AI involvement in the content. To determine when to apply AI labels, we use a multi-pronged approach:

---

<sup>55</sup> Meta, '[Meta Ad library](#)', May 2026

<sup>56</sup> Meta, '[Expanding GenAI Transparency for Meta's Ads Products](#)', Meta Newsroom, February 2025

- If an advertiser is using our in-house generative AI creative features and these tools do not result in significant edits to the image or video and do not include a photorealistic human, then we will not apply any AI labels.
- If an advertiser is using our in-house generative AI creative features and these tools result in a significant edit to the image or video, we will apply an AI label either behind the three-dot menu, or next to the Sponsored label. When these tools result in the inclusion of an AI-generated photorealistic human, the label will appear next to the Sponsored label (not behind the three-dot menu).

Meta also continues to maintain heightened disclosure requirements for political and social issue advertising. Our frameworks in these areas prioritise transparency and limit the potential for AI-generated content to mislead voters. Since 2024, advertisers running social issue, elections, and politics ads are required to disclose when their ad contains a photorealistic image or video, or realistic sounding audio, that was digitally created or altered by AI or other methods 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.

Advertisers running these ads do not need to disclose when content is digitally created or altered in ways that are inconsequential or immaterial to the claim, assertion, or issue raised in the ad. This may include image size adjusting, cropping an image, color correction, or image sharpening, unless such changes are consequential or material to the claim, assertion, or issue raised in the ad.

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.

We will continue to evolve our approach to labeling AI-generated content in partnership with experts, advertisers, policy stakeholders and industry partners as people's expectations and the technology change.

## **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.**

- **Commitment 35: Meta will continue to support research and events in relation to misinformation and media literacy.**

Since the 2025 transparency report, we have supported a number of initiatives on misinformation and media literacy, including: commissioning research from Associate Professor Timothy Graham at Queensland University of Technology on community-based content moderation, examining how Australians perceive and engage with participatory content moderation systems and the factors that drive or inhibit community participation, which will be published in 2026; and partnering with the Australian Associated Press FactCheck to develop and deliver a new AI literacy video campaign aimed at equipping Australians with the critical thinking skills needed to identify and evaluate AI-generated content online.

In addition, ahead of the 2025 federal election, Meta partnered with AAP to deliver a media literacy education campaign highlighting the importance of critical thinking skills for evaluating online content. Available in English, Simplified Chinese, Vietnamese and Arabic, the campaign reached 823,638 people in Australia, was viewed over 5.07 million times, and generated more than 2.45 million ThruPlays.<sup>57</sup>

Meta also supported ‘Facebook and the News Media: How Australians Engage with News and Misinformation Online,’ led by La Trobe University’s Professor Andrea Carson and the University of Waikato’s Dr. Justin Phillips, with Ms. Phoebe Hayman.<sup>58</sup> Published in September 2025, the report explored the evolving relationship between Australian news organisations, Facebook, and online audiences by analysing over three million posts from 25 major Australian news publishers spanning 15 years. The data was accessed by the researchers via the Meta Content Library. Notably, the study revealed that arts and lifestyle posts now dominate user interaction on Facebook, reversing a decade-old pattern where political stories led engagement.

Meta also convened a hybrid roundtable in Sydney in April 2025, bringing together government, regulatory, academic, and industry stakeholders to discuss information integrity ahead of the federal election. This session, ‘Supporting Media Literacy and Information Integrity in the AI Age’, explored how to mitigate AI-driven disinformation risks and strengthen media literacy as generative AI becomes increasingly sophisticated.

- **Commitment 36: Meta will continue to support research and events in relation to disinformation.**

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 also co-hosted with DIGI the annual Australian Code of Practice on Disinformation and Misinformation event in September 2025, ‘Information Integrity in the Gen AI Era’, which convened more than 50 participants from government, academia, civil society, and industry. Discussions focused on the disinformation challenges and opportunities arising from the proliferation of generative AI.

---

<sup>57</sup> Campaign results supplied by Australian Associated Press.

<sup>58</sup> Andrea Carson, [‘Facebook and the News Media: How Australians Engage with News and Misinformation Online’](#), September 2025

- **Commitment 37: Meta makes the Content Library and Content Library API services available to third-party fact-checking partners and eligible academics.**

Meta provides research tools called Content Library and Content Library API that share publicly accessible data from Facebook and Instagram. The Content Library and Content Library API research tools are available to qualified stakeholders.

- **Commitment 38: Meta will continue to support new areas of research relating to misinformation and disinformation in 2025.**

Please see research supported in 2025 outlined above.

## **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**

- **Commitment 39: 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.**

The publication of this report satisfies this requirement for 2026.

## Appendix A: 2025 specific commitments made by Meta under the industry code on disinformation and misinformation

The 39 commitments for 2025 are outlined below:

<b>Outcome 1: Combatting misinformation / disinformation</b>	<b>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.</b> <ol style="list-style-type: none"><li>1. Meta removes networks of accounts, Pages and groups that violate our Inauthentic Behaviour (IB) policy, including disinformation that violates our policy on Coordinated Inauthentic Behaviour (CIB).</li><li>2. Meta provides transparency about accounts, Pages and groups removed under our Inauthentic Behaviour policy.</li><li>3. Meta partners with experts and organisations who assist in providing tips or further investigation about possible inauthentic behaviour on our services.</li><li>4. Meta removes misinformation that violates the misinformation section of our Community Standards.</li><li>5. Meta removes misinformation that violates our policy against posting content that directly contributes to a risk of interference with participation in election processes.</li><li>6. Meta removes content that violates our policies prohibiting Coordination of Harm and Promoting Crime that calls for violence, coordinated interference in elections or promotes crime.</li><li>7. Meta removes fake accounts.</li><li>8. Meta allows for appeals in instances where users may disagree with our enforcement, including to the independent and external Oversight Board.</li><li>9. Meta has partnered with third-party fact-checking organisations, in Australia, to assess the accuracy of content on our services.</li><li>10. Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta applies a warning label to content found to be false, partly false, altered, or missing context by third-party fact-checking organisations.</li><li>11. Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta reduces the distribution of content found to be false, partly false, or altered, by third-party fact-checking organisations.</li><li>12. Where the third party fact checking organisations are engaged to fact-check content on our services and do so, Meta proactively searches for content that</li></ol>
--	--

matches content debunked by our fact-checking partners, to apply the same treatments.

13. Meta takes action on Pages, groups, accounts, or websites found to repeatedly share misinformation.
14. Meta builds safeguards into our GenAI features and models so that people can have a safer and more enjoyable experience.
15. Meta provides tools to help people know when they see content that was made with AI.

**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.**

16. 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.**

17. Meta makes on-platform reporting channels available to users for false information.
18. Where the third party fact checking organisations are engaged to fact-check content on our services and do so, 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.**

19. Meta makes global transparency reports available regularly.
20. Where available and useful, 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.**

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

	<p>22. Meta takes steps to limit the possible distribution of misinformation via recommendations.</p>
<p><b>Outcome 2:</b> <b>Disrupt monetisation and advertising incentives</b></p>	<p>23. Meta sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation.</p> <p>24. Meta commits to prevent fact-checked content from being advertised or monetised.</p>
<p><b>Outcome 3:</b> <b>Combat inauthentic user behaviour</b></p>	<p>See items listed under Outcome 1.</p>
<p><b>Outcome 4:</b> <b>Empower consumers to be informed</b></p>	<p>25. Meta provides contextual information around posts that users see from public Pages.</p> <p>26. 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.</p> <p>27. Meta reduces the distribution and recommendation of problematic and low quality content on Facebook and Instagram.</p> <p>28. Meta commits to providing transparency in relation to AI-generated images, videos, and audio content, when we detect industry-standard AI image indicators or when users disclose that they're uploading AI-generated content.</p> <p>29. Meta will look for opportunities to continue to work with the Government on other ways to promote authoritative information.</p>
<p><b>Outcome 5:</b> <b>Political advertising</b></p>	<p>30. Meta requires all advertisers of political and social ads to complete an ad authorisation, which includes verifying the advertiser's identity.</p> <p>31. Meta requires political and social issue ads to include a disclaimer disclosing who is paying for the ad.</p> <p>32. Meta provides the Ad Library, a searchable archive of all political and social issue ads on our services in Australia.</p> <p>33. Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political and social issue ads.</p> <p>34. Meta provides transparency and information for users in relation to AI-generated content used in ads.</p>

**Outcome 6:  
Research**

- 35. Meta will continue to support research and events in relation to misinformation and media literacy.
- 36. Meta will continue to support research and events in relation to disinformation.
- 37. Meta makes the Content Library and Content Library API services available to eligible stakeholders.
- 38. Meta will continue to support new areas of research relating to misinformation and disinformation in 2025.

**Outcome 7:  
Annual reports**

- 39. 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.