Reset.
AUSTRALIA
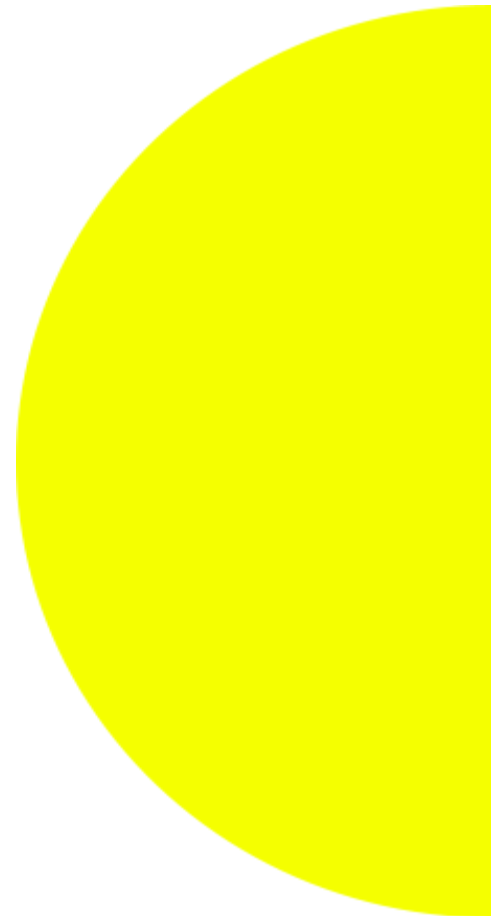
Response to Meta: Safeguarding the integrity of our Federal election

17 May 2022

On 2 May 2022, in collaboration with leading academics with expertise in mis- and disinformation, cyber abuse, and other online harms, Reset Australia sent Meta an open letter with 24 questions about safeguarding the integrity of our Federal election. On 6 May 2022, Josh Machin, Head of Public Policy, for Meta in Australia published answers to each of the questions in a public blog titled 'Update on Meta's work to support the 2022 Australian election'.

This is Reset Australia's response. In some instances, we have asked further questions, or clarifications, as many of the answers are inadequate or incomplete.

Question and Meta Answer	Reset Australia Response and Clarification
Preamble	
Meta has had a comprehensive strategy in place over the course of the Australian election campaign to combat misinformation, voter interference and potentially harmful content on our platforms. We shared details about our plans in a March 2022 blog post, and numerous roundtables with journalists, government agencies, law enforcement and security agencies (including via the Australian Government's election integrity assurance taskforce), and academics.	We welcome the opportunity for factual clarification of the public record, and for this transparent and robust public exchange regarding an issue that is fundamental to the functioning of our democracy: Meta's impact on the integrity of our Federal election. The public deserves much more detail and depth than was provided in your March 2022 blog post (titled 'How Meta is preparing for the 2022 Australian election'). In the absence of mandatory transparency measures through binding regulation, untrustworthy companies such as Meta require their behaviour to be closely scrutinised - particularly during crucial moments such as the final week of the election campaign.
In particular, we appreciate the close working relationship with the Australian Electoral Commission who has been referring content to us that they believe may violate Australian electoral law or represent voter interference. We are also continuing our collaboration with civil society organisations such as the Australian Strategic Policy Institute (ASPI) and First Draft.	Our intention in writing to you was to enable a broader set of stakeholders to learn more about the details of your election plans, including the general public, civil society organisations that do not have contractual arrangements with you (as is the case with ASPI[1] and First Draft), and a wider group of academics and experts (including the signatories to the original letter).
This week, we and other tech companies received further questions from Australian academics – coordinated by a technology-focussed lobby group – regarding the steps we are taking in the months before the Australian election. We agree that it is important for digital platforms to be transparent and accountable to the Australian public. In that vein, we are publishing responses to the questions below.	As stated in our original letter to you, "adequate regulatory frameworks are not yet in place" to holistically and systematically address the spread of

---

[1] Meta/Facebook is the largest private sector donor of the Australian Strategic Policy Institute (ASPI), providing $269,574 of funding to the organisation in 2020-21 [Reference: 2020-2021 Annual Report, Australian Strategic Policy Institute, available at: https://ad-aspi.s3.ap-southeast-2.amazonaws.com/2022-02/ASPI%20Annual-Report_2020-2021.pdf?VersionId=B7LG1EptgVEw39hBwyXHrxAzLueB3BSL]

Question and Meta Answer	Reset Australia Response and Clarification
We will have continued transparency of our efforts to combat misinformation in particular in our next annual transparency report, required under the voluntary industry code on misinformation and disinformation, which is due in May 2022. Although the reporting period covers 2021, we recognise the level of interest in the steps we are taking to promote integrity of the Australian election, so we will be voluntarily including further information in that report. You can find our previous report here. We remain committed to engaging with Australian academics and experts on these important policy issues.	mis- and disinformation and hate speech on digital platforms. Your response fails to acknowledge that the Australian Code of Practice on Misinformation and Disinformation, drafted and administered by industry group DIGI (of which you are a founding member) has been critiqued for a variety of reasons[2]. The March 2022 report by the House Select Committee on Social Media and Online Safety[3] and June 2021 Australian Communication and Media Authority (ACMA) report which assessed the Code[4] point to the need for stronger, more systematic regulation of our digital information ecosystem. Given the original ACMA position paper for the Australian Code[5] drew heavily on the EU Code of Practice on Disinformation, it is significant that just recently, in April 2022, the landmark EU legislation, the Digital Services Act (DSA)[6] (companion to the Digital Markets Act (DMA)) was approved. The DSA represents a paradigm shift in tech regulation as it sets rules and standards for algorithmic systems in digital media markets. It requires greater transparency about platforms' data and algorithms, including audits and fines of up to six percent of their annual sales for

[2] The draft Code released by DIGI was critiqued by ACMA as being "a long way from the model that we proposed to address these important issues", and recently described the Code as being "too narrow" to prevent the harms of mis- and disinformation [references: Samios, Z, & Vistenin, L. (2020) ACMA: Tech giants' code to handle fake news fails to meet expectations. Available at: https://www.smh.com.au/politics/federal/acma-tech-giants-code-to-handle-fake-news-fails-to-meet-expectations-20201026-p568oq.html; Karp, P (2022) Digital code of conduct fails to stop all harms of misinformation, Acma warns. Available at: https://www.theguardian.com/media/2022/mar/21/digital-code-of-conduct-fails-to-stop-all-harms-of-misinformation-acma-warns]. Reset Australia and other digital rights organisations have critiqued the self-regulatory model of the Code that means it is voluntary, opt-in, has little enforcement mechanisms and no penalties [reference: Smith, P. & Ward, M. (2022) Tech giants' 'laughable' disinformation solution slammed. Available at: https://www.afr.com/technology/tech-giants-laughable-disinformation-solution-slammed-20211010-p58ys8].

[3] Parliament of the Commonwealth of Australia (2022) Social Media and Online Safety. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024877/toc_pdf/SocialMediaandOnlineSafety.pdf;fileType=application%2Fpdf

[4] Australian Communication and Media Authority (2021) Report to government on the adequacy of digital platforms' disinformation and news quality measures. Available at: https://www.acma.gov.au/sites/default/files/2021-11/Adequacy%20of%20digital%20platforms%20disinformation%20and%20news%20quality%20measures.pdf

[5] Australian Communication and Media Authority (2020) Online misinformation and news quality in Australia: Position paper to guide code development. Available at: https://www.acma.gov.au/online-misinformation-and-news-quality-australia-position-paper-guide-code-development

[6] European Commission (2020) Proposal For A Regulation of the European Parliament & of the Council on a Single Market For Digital Services (Digital Services Act) & Amending Directive. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en

Question and Meta Answer	Reset Australia Response and Clarification
	repeated infringements. Post-election, Australia has an opportunity to reduce this widening gap between our regulatory framework in Australia and the EU and consider which DSA elements could be applicable in our context (refer to Reset Australia policy brief titled 'The future of digital regulation in Australia: Five policy principles for a safer digital world' for further detail regarding aspects of the DSA that are relevant to Australia[7]).
	Given the inadequacies of the current Code, Meta's reporting against it in the recent transparency report does not assist us with the task of evaluating the efficacy of Meta's election safeguards. In the June 2021 report[8] ACMA stated that signatories "lacked systematic data, metrics or key performance indicators (KPIs) that establish a baseline and enable the tracking of platform and industry performance against code outcomes over time". An example of this is Meta's statistic that it removed 14 million pieces of COVID-19 misinformation content between March 2020 and December 2020. This data is not comparative or success-oriented and hence gives no indication of the effectiveness of Meta's content moderation systems.
	Further to this, as former Facebook executive and whistleblower Frances Haugen stated in a testimony before the House Select Committee on Social Media and Online Safety, social media giants have kept the online safety discourse focused on content moderation systems that deal with harmful and illegal content downstream, rather than directing attention upstream to the algorithms that amplify this content in the first instance. These algorithms "have so much

---

[7] Reset Australia (2022) The future of digital regulation in Australia: Five policy principles for a safer digital world. Available at: https://au.reset.tech/uploads/the-future-of-digital-regulations-in-australia.pdf

[8] Australian Communication and Media Authority (2021) Report to government on the adequacy of digital platforms' disinformation and news quality measures. Available at: https://www.acma.gov.au/sites/default/files/2021-11/Adequacy%20of%20digital%20platforms%20disinformation%20and%20news%20quality%20measures.pdf

Question and Meta Answer	Reset Australia Response and Clarification
	sway over our democratic outcomes", she warned[9]. Meta's recent transparency report makes no mention of how harmful and misleading content is amplified by platform algorithms.

1. How many dedicated human content moderators will be bolstering your AI enabled system specifically for the Australian election?

In addition to the significant investments Meta has made in technology to proactively identify harmful content, we also have more than 40,000 people who work on safety and security at Meta (about 15,000 of which are dedicated content reviewers). As issues potentially arise during the Australian election, we have the benefit of being able to draw from not just the cross-functional team dedicated to the Australian election but also from any members of our global safety and security teams as needed, depending on the expertise and skills required. Another benefit of investing so heavily in safety and security means that we have 24/7 support; even while Australia sleeps, our safety and security teams in other timezones are able to review content that could be harmful and impact the Australian election.	The March 2022 blog post also stated that Meta has '40,000 people around the world working on safety and security'. It is helpful to know that 15,000 people are dedicated content reviewers. We also note that your submission to the First Interim Report of the Senate Select Committee on Foreign Interference through Social Media states that you have more than 35,000 people who 'work with technology to apply their own experience and knowledge to detect and assess possible networks of CIB (coordinated inauthentic behaviour)'[10].
	Are you able to provide clarification of the above as follows:
	1. As per the original questions, what is the size of the specific 'cross functional team dedicated to the Australian election'?
	2. Are the 40,000 people working on safety and security across the Meta businesses i.e. across Facebook, Whatsapp, Messenger and Instagram?
	3. Are the 15,000 dedicated content reviewers responsible for reviewing content across Facebook, Whatsapp and Instagram globally (in at least 150 countries[11])?

---

[9] Parliament of the Commonwealth of Australia (2022) Social Media and Online Safety Public Hearings. Available at:
https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Media_and_Online_Safety/SocialMediaandSafety/Public_Hearings
[10] Parliament of the Commonwealth of Australia (2021) Foreign Interference through Social Media. Available at:
https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024741/toc_pdf/FirstInterimReport.pdf;fileType=application%2Fpdf
[11] Meta (2022) Launching Facebook Reels Globally and New Ways for Creators to Make Money, Available at:
https://about.fb.com/news/2022/02/launching-facebook-reels-globally/

Question and Meta Answer	Reset Australia Response and Clarification
	4. Given Meta's Daily Active People (DAP)[12], was 2.87 billion on average for March 2022[13], is 15,000 dedicated content reviewers sufficient? 5. Are the 35,000 people working to detect and assess CIB (as described in the First Interim Report of the Senate Select Committee on Foreign Interference through Social Media) part of the 40,000 people working on safety and security?
2. What languages do these content moderators speak? (for instance, the top five spoken languages in Australian homes other than English are Cantonese, Mandarin, Italian, Arabic and Greek)	
The benefit of drawing from our global investment in safety and security means that Meta supports content moderation in over 70 languages. We also have a global network of more than 80 third-party fact-checkers who cover more than 60 different languages to combat misinformation. We have run a consumer awareness campaign prior to the Australian election to help Australians spot misinformation and raise awareness of fact-checking. As well as running this campaign in English, we have translated campaign assets into Chinese, Vietnamese and Arabic (the top languages other than English spoken at home in Australia according to the Australian Bureau of Statistics). In the lead-up to the election, we have received feedback from non-government experts that many non-English speaking diaspora communities in Australia may also use digital platforms from their home countries (see here for example). We continue to encourage policymakers to pay regard to the risks of misinformation and disinformation that can occur in Australian communities on non-English platforms.	Thank you for your response. It is encouraging to hear that your content moderators globally speak over 70 languages and that AI technology detects hate speech in over 50 languages. 1. Can you confirm that this is the number of languages that will be moderated in the current Australian Federal election specifically? Given the scale and scope of Meta's operation globally it is difficult to discern the level of local investment here in Australia when global statistics and information is provided by your company. You have mentioned a global network of 80 third-party fact checkers working in 60 different languages, however you also state that your fact-checking partners for the Australian Federal election are primarily AFP, AAP and RMIT FactLab. 2. Are you able to clarify what languages the teams at AFP, AAP and RMIT FactLab are undertaking fact-checking in? (Note: we have

---

[12] Daily Active People or DAP is the number of people who have logged into Facebook, Instagram, Messenger, and/or WhatsApp
[13] Meta (2022) Meta Reports First Quarter 2022 Results, Available at:
https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-First-Quarter-2022-Results/default.aspx

Question and Meta Answer	Reset Australia Response and Clarification
	also reached out to these teams directly for a response to this). The concerns you raise about non-English speaking diaspora communities in Australia using so-called 'non-English platforms' is not relevant to the question. A large proportion of non-English speaking people use your platforms (a recent study estimated that 49.6% of Facebook users are non-English speaking[14]), including within Australia, hence the distinction between 'English' and 'non-English' platforms is not useful. Reset Australia scrutinises the behaviour of all major social media platforms, however the focus here is Meta. Your platforms play an outsized role in the Australian election with Facebook being Australia's most popular social media platform for news[15].

3. Where are the content moderators dedicated to the Australian election based? If not in Australia, can their security and integrity be ensured during this time of geo-political instability?

Our safety and security teams are located around the world, with centres of excellence in 20 sites around the world, such as Singapore, Dublin and the United States, which ensures around-the-clock coverage.	As mentioned in the previous question, it is difficult to discern the level of local investment here in Australia when global statistics and information are provided by your company. 1. Can you confirm that content moderators across 20 sites around the world are specifically dedicated to the Australian election, even during this time of geo-political instability where there is, in the words of Communications Minister Paul Fletcher, "a significant volume of content promoting violence, extremism and disinformation in relation to the Russian invasion of Ukraine"[16]?

[14] We Are Social (2021) Digital 2021. Available at: https://wearesocial.com/uk/blog/2021/01/digital-2021-uk/
[15] Reuters Institute & University of Oxford (2020) Digital News Report. Available at: https://www.digitalnewsreport.org/survey/2020/
[16] Hurst, D. & Butler, J (2022) Morrison government asks Facebook, Twitter and Google to block Russian state media 'disinformation'. Available at:
https://www.theguardian.com/world/2022/mar/03/morrison-government-asks-facebook-twitter-and-google-to-block-russian-state-media-disinformation

Question and Meta Answer	Reset Australia Response and Clarification

4. How has your content moderation system taken into account the nuances of Australian English slang? (Using Australian slang is a common strategy for those seeking to evade detection by content moderation system on social media.)

This is an important imperative not just in the context of the Australian election. We have dedicated teams with deep knowledge and expertise in Australia, including the use of language in the Australian context. These teams are trained to be across slang words and other colloquialisms. We always welcome input from local experts and civil society organisations about trends in possible abuse of our platform, and any academics are welcome to raise examples with us if they believe there is a new or emerging slang which is not properly being accounted for.	Thank you for acknowledging the importance of accounting for the nuances of Australian English slang. It is helpful to know that your teams are trained in slang words and other colloquialisms. Given your reliance on automated content moderation, alongside human content moderation (as described in the recent House Select Committee on Social Media and Online Safety[17]), 1. how do these technologies account for slang and colloquialisms? One example of slang not being accounted for is a recent experiment by Reset Australia[18] which sought to test Facebook's ad review system by attempting to approve advertisements that explicitly promoted election disinformation narratives that were common in the last US election. All five ads were approved by Facebook's ad platform, including one that used the term 'jab'. This indicates a need for greater transparency regarding the error rates of the automated content moderation systems (as per Question 6).

5. Who has been consulted in the development of election-related content moderation policies? How will you ensure these policies are adaptive and responsive to the events of the election?

Our election-related policies are global, and we have developed them through our involvement in over 200 elections around the world since 2017. We have engaged extensively with Australian stakeholders about these policies, especially those that have been developed since the 2019 Australian election like our voter interference policies, in the months leading up to	It is encouraging to hear that you have engaged extensively with Australian stakeholders about your company's election-related policies.

---

[17] Parliament of the Commonwealth of Australia (2022) Social Media and Online Safety. Available at:
https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024877/toc_pdf/SocialMediaandOnlineSafety.pdf;fileType=application%2Fpdf
[18] Reset Australia (2022) Facebook still approving ads with explicit disinformation as election campaigning
ramps up. Available at: https://au.reset.tech/uploads/facebook-electoral-disinfo-ad-experiment-1.pdf

Question and Meta Answer	Reset Australia Response and Clarification
the Australian election campaign – including providing transparency in <u>our last report</u> under the voluntary industry code on misinformation and disinformation. No stakeholder has raised concerns with us to date that these policies appear to be inadequate in the Australian context.	1. As the original question states, are you able to give an indication of who these Australian stakeholders are? Your last transparency report under the voluntary industry code on mis- and disinformation does not list specific civil society organisations (with the exception of UNICEF and Save the Children who are listed as authoritative sources for misinformation regarding COVID-19 and QAnon respectively). It is crucial for Meta to be transparent with the public regarding the expertise drawn upon, including that of communities most impacted by mis- and disinformation and hate speech, to develop election-related policies. Given the proliferation of anti-trans hate speech and disinformation on Facebook during this election campaign (as outlined in Reset Australia's Election Radar[19] on 'Anti-trans hate speech and misinformation surges on social media as Deves' campaign occupies headlines'), 2. have you actively engaged with civil society organisations with expertise that could support your response and management of this, and the risks it poses to the wellbeing and safety of trans and gender-diverse people, particularly children and young people? Legal action launched by the Australian Muslim Advocacy Network (AMAN) and the Islamic Council of Queensland against former Senator Fraser Anning with regards to content posted on Facebook and Twitter in 2021 in the Queensland Civil and Administrative Tribunal found his posts breached state laws that ban the incitement of hatred, serious contempt or severe ridicule on the grounds of

---

[19] Reset Australia (2022) Anti-trans hate speech and misinformation surges on social media as Deves' campaign occupies headlines. Available at: https://au.reset.tech/news/election-radar-anti-trans-hate-speech-and-misinformation-surges-on-social-media-as-deves-campaign-occupies-headlines/

Question and Meta Answer	Reset Australia Response and Clarification
	religion.[20] Mr Anning was ordered to remove 141 pieces of content but still maintains a presence on Facebook despite being a known source of misinformation about COVID-19 vaccines. 3. What civil society organisation and other experts have you engaged with to support your response to disinformation targeting Muslims and ethnicities associated with Islam? 4. What measures have you put in place to address hate speech and disinformation targeting Muslims and ethnicities associated with Islam?
6. What AI enabled content moderation system will be deployed during the election, including image recognition technology? What are the error rates?	
We provide transparency about our content enforcement approach in our global transparency centre.	Meta's global transparency centre does not appear to provide error rates for image recognition technology. Both the Senate Select Committee on Foreign Interference through Social Media Interim Report[21] (refer to sections 3.26 - section 3.32) and the House Select Committee on Social Media and Online Safety Report (refer to section 3.46 to section 3.55)[22] detail concerns about the error rate of automated content moderation technologies, and the lack of transparency over these systems. Hence, further information would be welcomed by the Australian public.
7. What provisions have been made to protect communities from foreign interference?	
Foreign interference in Australia can occur via a variety of means. In relation to our apps, we take extensive	Beyond the steps taken to remove CIB and fake accounts, and the existing ad authorisation process

---

[20] Chalmers, M & Robertson J (2021) Fraser Anning ordered to remove Facebook and Twitter posts that tribunal found vilified Muslims. Available at: https://www.abc.net.au/news/2021-08-03/tribunal-orders-fraser-anning-to-remove-posts-vilifying-muslims/100337536
[21] Parliament of the Commonwealth of Australia (2021) Foreign Interference through Social Media. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024741/toc_pdf/FirstInterimReport.pdf;fileType=application%2Fpdf
[22] Parliament of the Commonwealth of Australia (2022) Social Media and Online Safety. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024877/toc_pdf/SocialMediaandOnlineSafety.pdf;fileType=application%2Fpdf

Question and Meta Answer	Reset Australia Response and Clarification
steps to identify and take action against threats to the election, including signs of coordinated inauthentic behaviour and block millions of fake accounts everyday so they can't spread misinformation. These are outlined in detail in our last transparency report. We also require all advertisers looking to run political, social and election related ads to complete an authorisation process, confirming their identification and be located in Australia. In particular, we continue to engage closely with Australian security agencies regarding the overall threat environment for foreign interference in Australia. We continue to monitor this closely over the final weeks of the election campaign.	you have had in place for electoral and political ads since August 2020 (which was subsequently expanded to social issue ads in June 2021), 1. what specific election related measures have you put in place with regard to protecting communities from foreign interference? In December 2021, the Senate Select Committee on Foreign Interference through Social Media Interim Report recommended that "the Australian Government take a proactive approach to protecting groups that are common targets of foreign interference but are not classified as government institutions"[23]. 2. Do any of the election preparations you have undertaken align with this recommendation?

8. What avenues are in place to enable civil society organisations to flag harmful and false content (beyond the reporting mechanism of the eSafety Commissioner)?

The community has a wide-array of mechanisms available to report potentially harmful or false content to us. We have easy and simple in-app reporting for every piece of content. We also have close working relationships with Australian regulators, like the eSafety Commissioner and Australian Electoral Commission, who are able to quickly and easily refer content to us. We take a 'no closed door' approach to content review where we will review any and all content sent to us by an Australian regulator. We also have built direct relationships with expert civil society organisations who are able to refer content to us directly.	Thank you for this response. Australian Muslim Advocacy Network has routinely submitted reports about false content to Meta and never received a response or communication of any outcome.[24] As stated in Question 5 above, 1. are you able to provide a list of expert civil society organisations you work with directly?

[23] Parliament of the Commonwealth of Australia (2021) Foreign Interference through Social Media. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024741/toc_pdf/FirstInterimReport.pdf;fileType=application%2Fpdf
[24] Parliament of the Commonwealth of Australia (2022) Social Media and Online Safety. Available at: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024877/toc_pdf/SocialMediaandOnlineSafety.pdf;fileType=application%2Fpdf

Question and Meta Answer	Reset Australia Response and Clarification
The industry association DIGI maintains a complaints mechanism for any instance of possible breaches of the industry code on misinformation and disinformation. Finally, for content on our services, any member of the community is able to raise it directly with our third-party fact-checkers for possible fact-checking: ● AFP: https://factcheck.afp.com/contact ● AAP: 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	

| **9. In the United States you have shared data about removed coordinated inauthentic behaviour networks with independent researchers. Why have you not implemented this in Australia? (ideally all content and accounts removed under election-related policies should be stored and shared for post-election scrutiny)** ||

| The assertion in this question is not correct. In late 2020, we launched a pilot, CrowdTangle-enabled, research archive where we've shared over 100 of the recent coordinated inauthentic behaviour (CIB) takedowns with a small group of researchers who study and counter influence operations. The Australian Strategic Policy Institute (ASPI) is one of our initial 5 key partners for this archive globally, where they can study and analyse the networks we've removed. We are pleased to see Australian representation in the small number of research organisations who are suitable for the pilot. | The pilot referred to is a global report on CIB and cyber espionage with no mention of Australia. In reviewing Meta's reports on CIB takedown[25], other regions of the world and countries are the focus, not Australia.

1. Will the next Quarterly Adversarial Threat Report include a focus on the Australian Federal election?

2. In the interests of greater transparency, are there plans to make this data available to a broader set of research organisations in Australia? |

---

[25] Meta (2022) Meta reports on CIB takedown. Available at: https://about.fb.com/news/tag/coordinated-inauthentic-behavior/

Question and Meta Answer	Reset Australia Response and Clarification
10. During the last United States election you labelled posts that were believed to be state-controlled media outlets. Why have you not implemented this in Australia?	
The assertion in this question is not correct. We label state-controlled media outlets in many countries around the world, including in Australia.	Thank you for this response. 1. Does this include pages or groups that are run by nation states?
11. Given that over 20% of Australians speak a language other than English at home, what languages will the third-party fact checks be translated into?	
We also have a global network of more than 80 third-party fact-checkers who cover more than 60 different languages to combat misinformation. Fact-checks are available in the language of the original content. A piece of content in a non-English language that arises in Australia is eligible to be fact-checked by other global partners who operate in that language.	As per Question 2, 1. are you able to clarify what languages the teams at AFP, AAP and RMIT FactLab are undertaking fact-checking in?
12. What non-English language publications will third-party fact checks be provided to?	
It's not clear what is meant by this question. More information about the scope and eligibility of third-party fact checking on our services is available at our Help Centre	Google collaborates with AAP to provide and translate fact-checks to 40 culturally and linguistically diverse publications free of charge (including: Koori Mail, Indian Link, Viet News, Epoch Times, Ngaarda Radio, Korean Herald, Australian Muslim Times, Phil Times, Almestaqbal and SBS' channels and platforms). 1. Are you able to provide a comparable figure?
13. How will the speed of fact checking be measured during the election?	
It's not possible to give a timeframe around how long it takes a fact checker to verify content after it is posted on Facebook. This is because content is flagged to fact checkers in a variety of ways, and it is at the discretion of the independent fact checkers as to which pieces of content they review. The amount of time it takes a fact checker to verify a claim and undertake a fact check	In the modern election contest, speed of fact checking is an essential success measure, not only due to the 24 hour news cycle, the fast moving political agenda, but most of all due to the potential for mis- and disinformation to spread virally on social media. It is understandable that the time taken by a fact checker may vary depending on the claim and its

Question and Meta Answer	Reset Australia Response and Clarification
can also vary, depending on the complexity of the claim they are reviewing. Content is flagged to our third-party fact-checkers to review in several ways: ● our third-party fact-checkers proactively identify the content themselves ● our technology identifies potential false stories for third-party fact-checkers to review. For example, when people on Facebook submit feedback about a story being false or comment on an article expressing disbelief, these are signals that a story should be reviewed ● we also have a similarity detection system that helps us identify more debunked content than what our fact checkers see. What's important is once a story is debunked by our third party fact checkers, we make it less visible on Facebook and Instagram. Artificial intelligence plays an important role in helping scale the efforts of our third party fact checkers. Once fact-checkers deem a piece of content is false and mark it as false, the overlay will appear almost immediately. After one fact check on one piece of content, we're able to kick off similarity detection which helps us identify duplicates of debunked stories, and reduce their distribution. These new posts are then fed back into the machine learning model which helps improve its accuracy and speed.	complexity, however, what we are seeking here is an approximate range and/or the average time taken. 1. Is debunked content, or content that has been deemed false through fact checking, taken down from your platform if it violates platform policies (for example, Facebook's COVID-19 policy[26])?

14. How will the reach of fact checking be measured during the election (i.e. how many people have viewed the content and their demographics)?

---

[26] Facebook (2022) COVID-19 and Vaccine Policy Updates & Protections. Available at: https://www.facebook.com/help/230764881494641/

Question and Meta Answer	Reset Australia Response and Clarification
Once a piece of content is found to be false, we apply a warning label on it so it is not possible to see the content without clicking past the warning label. Once a warning label is applied, 95% of people on Facebook choose not to click through. This dramatically reduces the number of people who see the content.	1. Are you able to provide more context around this statistic? For instance, what is the click through rate of content without a warning label? The intent of this question was to also understand the reach of the fact checking outputs (not only the content subject to fact checking). For example, RMIT FactLab publishes their fact checks in their CheckMate newsletter. 2. What is the scale and reach of this into various audience segments? 3. Is it reaching target audiences (i.e. those who are viewing the false content)? This is a crucial metric for understanding the success of fact checking as a tactic for combating mis- and disinformation.

15. How will your response to fact checking be measured during the election? (i.e. content take down, or reporting to the appropriate authority)

Our approach to combating misinformation is comprehensive and is broader than simply content takedowns or third-party fact checking. Meta has led the industry in terms of transparency under the voluntary industry code on disinformation and misinformation and we continue to look for opportunities about what integrity data we might possibly be able to make available after the election in the interests of transparency. We have expanded our third-party fact-checking program in Australia to include RMIT FactLab, who are joining our existing partners Agence France Presse and Australian Associated Press. We have also one-off grants to all our fact-checkers to increase their capacity in the lead up to the election.	This response does not answer the question, and merely repeats information you have already provided. Your responses to Question 13 and 14 state that once a story is debunked by third party fact checkers, it is made less visible on Facebook and Instagram, and a warning label is applied to it. 1. Are there other responses in place, and how do you measure their success? 2. Is there data available on the number of users who are served false content prior to the completion of the fact checking process? Your claim that Meta has led the industry in terms of transparency under the voluntary code on disinformation and misinformation is not evidenced and difficult to believe in the context of breaking news

Question and Meta Answer	Reset Australia Response and Clarification
	about Facebook executives deliberately shielding documents showing they wanted to cause havoc to influence new media bargaining laws[27]. The facts emerging from the whistleblower complaint to the ACCC confirms that your company made a calculated decision to enact a widespread news outage that would have life-threatening consequences for Australians, to secure major amendments and concessions in the News Media Bargaining Code[28]. Contrary to your assertion, Meta is a company that will endanger the safety and wellbeing of Australians to evade transparency and binding regulation.

16. During the last United States election, Facebook's algorithm was adapted to reduce the distribution of sensational and misleading material, prioritising content from authoritative sources. Why has this measure not been implemented in Australia?

The assertion in the question is not correct. We provide transparency around how our ranking and recommendation algorithms work via our Content Distribution Guidelines and Recommendation Guidelines. As these policies outline, we reduce the distribution of sensational and misleading material at all times, not just in the lead-up to election campaigns. We are also taking steps to promote authoritative information about the election, by providing prompts in the Feed of every Australian to direct them to the Australian Electoral Commission's website.	It was widely reported, including by the New York Times,[29] that Facebook adapted its news feed algorithm change to lift news from authoritative outlets over hyper-partisan sources during the post-election period as part of Facebook's emergency "break glass" plan to combat misinformation. One of the most shocking claims by former Facebook executive turned whistleblower, Frances Haugen, is that Facebook's decision to prematurely turn off misinformation safeguards (i.e. a suite of "break glass" measures, including the algorithm adaptation) following the US election, contributed to the January 6 Capitol insurrection[30].

---

[27] Swan, D. (2022) Facebook executives 'hid misconduct' amid news ban. Available at:
https://www.theaustralian.com.au/business/technology/facebook-executives-hid-misconduct-amid-news-ban/news-story/26a0f7ac8f50641213199eee0d8114b6

[28] Hagey, K., Cherney, M. & Horwitz, J. (2022) Facebook Deliberately Caused Havoc in Australia to Influence New Law, Whistleblowers Say. Available at:
https://www.wsj.com/articles/facebook-deliberately-caused-havoc-in-australia-to-influence-new-law-whistleblowers-say-11651768302?st=uvvgri48znpp5zw&reflink=article_copyURL_share

[29] Roose, K. (2020) Facebook Deliberately Caused Havoc in Australia to Influence New Law, Whistleblowers Say. Available at:
https://www.nytimes.com/2020/12/16/technology/facebook-reverses-postelection-algorithm-changes-that-boosted-news-from-authoritative-sources.html

[30] ABC (2021) Facebook rushed to ditch misinformation safeguards after US election and fed Capitol riot, whistleblower claims. Available at:
https://www.abc.net.au/news/2021-10-04/facebook-whistleblower-on-capitol-riots-misinformation/100512688

Question and Meta Answer	Reset Australia Response and Clarification
	1. Is Australia's election a large enough commercial priority to your company to warrant the deployment of measures that have been proven to be effective in the United States and other jurisdictions?

17. During the last United States election, the distribution of live videos related to the election was limited. Why has this measure not been implemented in Australia?

While we learn lessons from each prior election, no two elections are the same. Working closely with elections authorities and trusted partners in each country, and evaluating the specific risks ahead of each election, we make determinations about which defences are most appropriate. In the lead-up to each election, we monitor the threats on our platform and respond accordingly. We have strong integrity measures to protect the abuse of products like Facebook Live at all times, not just for the Australian election. Our Community Standards and third-party fact checking initiatives apply equally to livestreaming as other types of content on our services.	We understand that you have a general Facebook Live policy[31] that seeks to ensure that this product is not misused or abused. This question is referring to additional temporary steps that Meta may choose to put in place as it did for the last United States election[32]. The limitation of the distribution of live videos (related to the election) was a "break glass" (or emergency measure) deployed in response to reports of inaccurate claims about the election. 1. Would Meta consider limiting the distribution of election-related live videos in response to mis- or disinformation during or post the Australian election?

18. Google has restricted the targeting for election ads in Australia. Has Meta considered this? If so, why has it not been implemented?

Digital platforms provide a range of different services, which may lead to different assessments about the best approach to integrity measures. At Meta, our approach is grounded in industry-leading transparency for political and social issue ads in Australia. We require these advertisers to go through an authorisation process, to add a disclaimer, and to agree to their ads appearing in the Ad Library for seven years after they run. We are committed to providing transparency of these ads that appear on our services.	This response does not answer the question. This question is about ad targeting, not about ad review systems or ad transparency. We are aware that following the 2016 and 2020 United States election controversies, Meta has taken steps to increase the transparency over how the company sells access to voters during elections. As mentioned in the question, Google only permits ad personalisation based on geographic location (except

---

[31] Facebook (2022) Facebook Live Policies. Available at: https://www.facebook.com/policies/live
[32] Facebook (2020) A Look at Facebook and US 2020 Elections. Available at:
https://about.fb.com/wp-content/uploads/2020/12/US-2020-Elections-Report.pdf

Question and Meta Answer	Reset Australia Response and Clarification
	radius around a location), age, and gender. The recently approved EU Digital Services Act bans online platforms from targeting ads to children, and prohibits targeting based on particular characteristics (notably religion, sexual orientation, ethnicity or political affiliation)[33].

Meta does not limit ad targeting, even during elections. You allow advertisers to buy access to users based on their location, age, gender, demographics, behaviours, interests, use of Meta products, and more. Combining these different micro-targeting categories allows Meta to sell an unprecedented level of access to Australian voters to political parties, candidates, lobby groups, and others seeking to influence the outcome of the election.

Your Ad Library also does not make any of this micro-targeting information available and only outlines top-level demographic information about what targeting is being sold.

1. Given widespread concern about how political, electoral and social ads are influencing Australian elections (particularly on platforms such as Facebook), is your lack of willingness to limit ad targeting an evasion of accountability? |
| **19. During the last United States election, Meta implemented changes to ensure fewer people saw social issue, electoral and political ads that had a "paid for by" disclaimer. Why has this measure not been implemented in Australia?** ||
| The assertion in this question is not correct. In response to community feedback, we have been running <u>tests</u> in a number of countries to show political content lower in people's Feeds. This applies to | Thank you for the link to the information about tests to show political content lower in peoples' Feeds. This |

[33] European Commission (2020) Proposal For A Regulation of the European Parliament & of the Council on a Single Market For Digital Services (Digital Services Act) & Amending Directive. Available at:
https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en

Question and Meta Answer	Reset Australia Response and Clarification
organic, non-paid content and was announced in the US after the last election. In 2021, we announced a new control feature that allows people to have more control over the ads they see on Facebook. This feature gives people a choice to see fewer social issues, electoral, and political ads with "Paid for by" disclaimers in Australia.	confirms our concern that the United States and other jurisdictions are prioritised in comparison to Australia. 1. Will Australia be included in these tests? 2. Does Meta take into consideration upcoming national elections when determining which jurisdictions should be included in such projects?

20. During the last United States election, the creation of new ads about social issues, elections or politics in the last few days of the election was blocked. Why has this measure not been implemented in Australia?

While we learn lessons from each prior election, no two elections are the same. Working closely with elections authorities and trusted partners in each country, and evaluating the specific risks ahead of each election, we make determinations about which defences are most appropriate. In the lead-up to each election, we monitor the threats on our platform and respond accordingly. The decision on whether Australia's blackout period for electoral ads should be extended to digital platforms is a choice for policymakers. We have consistently said over many years we support extending this requirement to digital platforms.	We agree that the blackout period for electoral ads should be extended to digital platforms, and that this is a matter for policymakers. We look forward to your support of those measures as part of a broader regulatory approach to ensuring the integrity of elections are not undermined by digital platforms.

21. What measures do you have in place to screen the placement of ads to ensure all political ads are properly identified and labelled by the advertiser?

We take a number of steps to detect ads that should be classified as political or social issue ads but have not been correctly categorised by the advertiser. We do not make information about these detection steps available, to avoid providing information to bad actors on how to evade our policies. This review process may include the specific components of an ad, such as images, video, text and targeting information, as well	As described in response to Question 4, a recent experiment by Reset Australia sought to understand the approval of ads with overt disinformation narratives that were common in the last US election. The experiment intentionally selected 'no' when prompted by Facebook to state whether or not these ads were about "social issues, elections or politics". All five ads were approved, which highlights a

Question and Meta Answer	Reset Australia Response and Clarification
as an ad's associated landing page or other destinations, among other information. Australian regulators and civil society are welcome to refer ads to us that they believe should be categorised as political or social issues ads and do not have this disclaimer.	concerning flaw in Facebook's ad review system.[34] In your response to that research, a Meta spokesperson said "these ads never went live, and therefore our full enforcement detection technology did not have an opportunity to pick up these ads". However, this assertion that there is an additional scan of approved ads after the ad goes live has been challenged by an earlier Reset Australia study. That experiment utilised a consenting control group of Australian Facebook users to serve ads with false information. These ads ran unimpeded for months on end[35]. Meta is yet to formally respond to this study. Given your supposed reliance on this secondary scan of approved ads that have gone live, are you able to provide further clarity regarding the following: 1. What public information is available about this additional ad detection mechanism, including the average time the process takes? 2. For users who are served false or misleading ads prior to detection by this secondary system, does Meta take steps to correct the record? 3. What is the proportion of ads that evade the first detection mechanism but are then caught in the second?
22. Beyond the Ad Library, will Meta be making available a comprehensive public archive of all sponsored political content, including targeting data and aggregated engagement statistics by target audiences (accessible by API)?	
Meta already makes available a comprehensive public archive of all political and social issue ads on our services, via the Ad Library. The Ad Library provides industry-leading transparency, including metrics such as estimated audience size, impressions, and statistics	We are familiar with the Ad Library, and asked this question specifically due to the lack of detailed data available, particularly in relation to how political,

[34] Wilson, C. (2022) Facebook approved five obviously fake Australian election ads. Can we trust them to police the poll? Available at:
https://www.crikey.com.au/2022/02/28/facebook-approved-obviously-fake-australian-election-ads/
[35] Reset Australia (2022) How tough is Facebook's misinformation crackdown? Available at:
https://au.reset.tech/news/how-tough-is-facebook-s-misinformation-crackdown/

Question and Meta Answer	Reset Australia Response and Clarification
of the target audience (such as location and gender of the audience). This is also already available via an API to approved academics and experts.	electoral and social issue ads are targeted to users. For example, we know that: ● Facebook does not provide information about the targeting of ads to researchers or users being targeted, only data on ad impressions. ● It only provides three demographic variables (state, gender and age range) alongside an estimation of how many accounts were in the target group and how many times an ad was displayed. This provides little insight into how ads are being targeted. The more specific targeting is, and the potentially more problematic, the less relevant the provided information is. ● Meta has much more available data about how ads are targeted, such as the targeting criteria selected by advertisers, though this is not provided. ● The state level of location granularity falls far behind what Google provides in its ad library. Google is providing ad impressions data by postcode; Meta only by state. Meta cannot describe the transparency provided by the Ad Library as "industry-leading". As the original question states: 1. Will Meta be making this more detailed data about all sponsored political content publicly available?
23. What 'break glass' (or emergency) measures are on standby during the election?	
We maintain a series of systems already to help protect the integrity of elections on our platforms. We continue to monitor threats on our platform and respond accordingly. At this stage, we have not been advised	According to Meta's report titled 'A Look at Facebook and US 2020 Elections', "scenario planning" was undertaken in advance "to ensure we could adapt to changing circumstances in real-time", including

Question and Meta Answer	Reset Australia Response and Clarification
by law enforcement or intelligence agencies that the risk of real-world harm is high.	developing "a range of capabilities" that could be implemented if necessary. 1. The public deserves to know whether a similar process of scenario planning has been undertaken for the Australian election? As you have mentioned in your responses, "no two elections are the same". Whilst the likelihood of an equivalent event to the US Capitol insurrection in Australia may not be high, it is clear that 'real-world harm' has taken place during the Australian election campaign, and the high risk continues as we enter the final week. The previously outlined example of the widespread anti-trans hate speech and misinformation[36] has harmed trans and gender diverse people (particularly children and young people), who advocacy groups have expressed lack adequate support services and experience high rates of suicide[37]. This is an emergency that social media giants, including Meta, have failed to act on. 2. Has Meta put in place measures to respond to the proliferation of anti-trans hate speech and disinformation on social media during this Federal election campaign?
24. What type of event (in terms of reach and impact) would trigger the implementation of 'break glass' measures?	
See above.	There are a wide range of emergency scenarios that could eventuate during an Australian election. Your refusal to answer this question suggests that either you have not considered these, or that you are unwilling to be transparent about the plans you have in place. This is disappointing as we head into the final days of the election, as well as the critical

---

[36] Reset Australia (2022) Anti-trans hate speech and misinformation surges on social media as Deves' campaign occupies headlines. Available at: https://au.reset.tech/news/election-radar-anti-trans-hate-speech-and-misinformation-surges-on-social-media-as-deves-campaign-occupies-headlines/

[37] Equality Australia (2022). "Lifesaving" – trans equality and clinicians groups defend trans healthcare. Available at: https://equalityaustralia.org.au/media-release-lifesaving-trans-equality-and-clinicians-groups-defend-trans-healthcare/

Question and Meta Answer	Reset Australia Response and Clarification
	post-election period - in which the movement that led to the Capitol insurrection mobilised.

Reset Australia's Election Radar[38] ('"We're Coming For You": Malcolm Robert's Viral Vaccine Conspiracy Speech Ignites Threats of Violence'), describes how on 29th March, a speech by One Nation Senator Malcolm Roberts alleging a conspiracy to cover-up the harms of Covid-19 vaccines gained significant traction across multiple social media platforms including Facebook (clearly in breach of Facebook's COVID-19 policy[39]) . Comments across these various posts have called for violence against politicians and public servants.

   1. Whilst this event may not be considered an emergency that warrants 'break glass' measures, is Meta monitoring and responding to events such as this? What does this response entail? |

---

[38] Reset Australia (2022) "We're Coming For You": Malcolm Robert's Viral Vaccine Conspiracy Speech Ignites Threats of Violence. Available at: https://au.reset.tech/news/election-radar-we-re-coming-for-you-malcolm-robert-s-viral-vaccine-conspiracy-speech-ignites-threats-of-violence/

[39] Facebook (2022) COVID-19 policy updates and protections. Available at: https://www.facebook.com/help/230764881494641/