# Reset.
## AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*

# An experimental evaluation using the Voice referendum

# *Summary*

There are significant gaps in Australia's regulatory framework when it comes to electoral misinformation and disinformation served through paid-for advertising. This creates notable vulnerabilities in Australia's online information architecture, such as the growing problem of threats to electoral integrity.

This small research piece demonstrates issues with platform responses to electoral misinformation served through paid-for advertising and weaknesses in platform transparency reports to the *Australian Code of Practice on Disinformation and Misinformation* (the Code).

We put forward a range of paid-for ads containing explicit electoral misinformation for approval to run on Facebook, TikTok and X(Twitter). We found the following:

- TikTok's system appeared to catch some political advertising and misinformation, but not the majority. We submitted ten ads containing paid-for misinformation to test TikTok's ad approval system, and 70% were approved. TikTok approved seven ads, rejected one ad and did not review the final two after detecting the violating ad.

- Facebook's system appeared entirely dependent on an advertiser's self-declarations regarding the nature of the advertising, which evidently offers insufficient protection against bad actors. We submitted twenty ads containing paid-for misinformation to test Meta's ad approval system, and 95% were approved. Meta approved all nineteen ads that were not self-identified as 'political ads', rejecting only one ad that we had voluntarily identified as a political ad.

- X's (Twitter's) system did not request self-identification for political ads, nor did their system detect or reject it. We submitted fifteen posts containing paid-for misinformation to test X's ad approval system, and 100% were approved and scheduled to run.

For ethical reasons, none of these ads were run, as we cancelled them after gaining approval. To be clear, no misinformation was published as a result of this experiment.

Each platform creates an annual report around their handling of political advertising and misinformation in their transparency reports under the voluntary Code. However, none of these reports adequately addressed these issues. This experiment suggests that self-reporting mechanisms under the Code may be weak and require more effective scrutiny.

It is simply too easy to propagate electoral misinformation via paid-for ads. Either platform policies are inadequate, or the ad approval systems deployed by platforms are not up to the task of accurately detecting misinformation. Legislators and regulators must consider a risk-based, independently assessable and more comprehensive approach to social media regulation to address these vulnerabilities.

# Reset.
## AUSTRALIA

Susan McKinnon Foundation

# *Table Of Contents*

# Introduction

Monetised electoral misinformation and disinformation, in the form of paid-for advertising, is a significant problem for democracies all around the world. Governments and digital platforms should have policies and processes in place to address this, and these should be effectively implemented. The purpose of this research was to test the efficacy of these policies and processes in the Australian context, using the Voice referendum as a case study.

The toxic impacts of misinformation and disinformation on democracies over the past decade have been well documented. Misinformation and disinformation is more than merely a speech issue; it is a phenomenon unique to the networked digital world involving specific actors, behaviours, content and distributive effects. With the rise of misinformation and disinformation production as a key campaigning tactic, attacks on electoral integrity, in particular, have been attractive for opportunistic actors. Attempts to sow voter mistrust in ballot measures and public institutions have sadly become an expected theme in the lead-up to elections and referendums. Rapid and sustained research and advocacy efforts from journalists, researchers and civil society on the harms suffered from digitally enabled attacks on electoral integrity contributed to a suite of expanded platform policies[1] and targeted public policy efforts[2] over the last few years. Expanded platform policies included specific efforts to safeguard electoral processes, such as civic integrity policies.

Failures in platform policies and processes can allow misinformation and disinformation to be propagated by paid-for advertising, presenting significant risks to Australia's political landscape.[3] Disinformation through advertising is one of the key tools available to the 'disinformation for hire' industry, such as those who engage in coordinated, inauthentic campaigns to influence electoral processes and debates. We have seen how Cambridge Analytica deployed 'weapons grade' data harvested from social media platforms and psycho-ops tactics using targeted advertising and disinformation on social media to interfere with elections in Trinidad and Tobago, Kenya, potentially the UK's Brexit vote and President Trump's campaign.[4]

---

**1.** For example, the lead-up to the U.S. midterms in 2022 culminated in expansions to Twitter's Civic Integrity Policy and refinements to Facebook's 'Voting Center': Naomi Nix, 2022 'Inside the civil rights campaign to get Big Tech to fight the 'big lie'', *The Washington Post*, 11th September 2022 https://www.washingtonpost.com/technology/2022/09/22/midterms-elections-social-media-civil-rights/.

**2.** See, for example: European Commission High Level Expert Group on Fake News and Online Disinformation 2018 'A multi-dimensional approach to disinformation - Report of the independent High level Group on fake news and online disinformation' *European Commission*, https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation; European Commission 2018 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Tackling online disinformation: a European Approach' (COM/2018/236 ) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236; European Parliament and the Council of the European Union 2022 'Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (*Digital Services Act*)' https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065

**3.** See Elise Thomas 2022 'Conspiracy Clickbait: Farming Facebook' *Institute for Strategic Dialogue* https://www.isdglobal.org/isd-publications/conspiracy-clickbait-farming-facebook/

**4.** Larry Madowo 2018 'How Cambridge Analytica poisoned Kenya's democracy' *Washington Post* https://www.washingtonpost.com/news/global-opinions/wp/2018/03/20/how-cambridge-analytica-poisoned-kenyas-democracy/, Eric Auchard 2018 'Cambridge Analytica Stage Managed Kenyan PResident's Campaign' Reuters https://www.reuters.com/article/us-facebook-cambridge-analytica-kenya-idUSKBN1GV300, Mark Scott 2019 'Cambridge Analytica did work for Brexit Groups' *Politico* https://www.politico.eu/article/cambridge-analytica-leave-eu-ukip-brexit-facebook/ , Peter Lewis & Paul Hilder 'Leaked Cambridge Analytica's Blueprint for the Trump Victory' *The Guardian* https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory

Reset. AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*    4

Or, more recently, Team Jorge – a 'black ops' disinformation unit – claims to have manipulated over thirty elections around the world using hacking and automated disinformation on social media from Nigeria to Kenya, as well as hijacking French news broadcasters to protect Russian interests in Monaco.[5] Both Cambridge Analytica and Team Jorge highlight the convergence of 'coordinated inauthentic behaviour' tactics and criminal activities, from data privacy breaches to bribery and hacking, with their criminal success ultimately resting on the failure of digital platforms to effectively protect against their activities.

The scale of the risks posed by allowing misinformation and disinformation to be propagated via paid-for advertising depends entirely on platforms' ability, or lack thereof, to efficiently implement effective policies and processes. Understanding the effectiveness of a platform's responses in Australia using the Voice referendum as a case study is critical for comprehending the safety and security of Australia's online information architecture. In this research, we stress test three platforms' policies and procedures: TikTok, Facebook and X (Twitter).

## Why should the Voice referendum be used as a case study?

The Voice referendum is a uniquely important event in Australia's history and provides a valuable, timely case study for evaluating platform responses to misinformation and disinformation, specifically the following:

- It is distinctly Australian, which means we can monitor international platforms' responses to an Australian issue, as there is less potential conflation with global responses.

- It is an Australian electoral process, meaning that all the features of electoral misinformation and disinformation will apply, and learnings can be made for future elections.

- To an extent, it is more narrowly defined than a broader election, where 'electoral content' and 'general current affairs content' can become harder to differentiate between.

5. Stephanie Kirchgaessner, Manisha Ganguly, David Pegg, Carole Cadwalladr and Jason Burke 2023 'Revealed: The hacking and disinformation team meddling in elections around the world' *The Guardian* https://www.theguardian.com/world/2023/feb/15/revealed-disinformation-team-jorge-claim-meddling-elections-tal-hanan

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid–for advertising?*
**An experimental evaluation using the Voice referendum**

5

# What should happen: Electoral law, the Digi Code & platform policies

There are significant gaps in Australia's regulatory framework when it comes to electoral misinformation and disinformation served through paid-for advertising. This creates vulnerabilities in Australia's online information architecture, leaving us exposed to attacks from bad actors.

- It is a criminal offence under Australian electoral law and referendum procedure to print, publish or distribute, or, *cause, permit, or authorise* the publication or distribution of anything likely to mislead or deceive an elector in relation to the casting of a vote.

- The *Australian Code of Practice on Disinformation and Misinformation* ('the Code')[6] provides unclear directions when it comes to electoral misinformation propagated by paid-for advertising. Misinformation propagated through paid-for advertising exists in a 'grey area'; it may be covered by the Code when it is propagated by inauthentic behaviour (automation, bots, etc.), but it may not be covered when it is propagated by human users.

- Platforms' policies and processes vary. All platforms have policies that should prevent electoral misinformation from being promoted in paid-for advertising and use automated or human moderation processes to detect this.

This set of overlapping and potentially contradictory policies means there is some ambiguity as to what the response to electoral misinformation and disinformation propagated in paid-for ads should be.

## Electoral law and referendum procedure

It is a criminal offence in s 329(1) of the *Electoral Act* 1918 to publish or distribute material likely to deceive an elector in relation to casting a vote.

> **A person shall not, during the relevant period in relation to an election under this Act, print, publish or distribute, or cause, permit or authorise to be printed, published or distributed, any matter or thing that is likely to mislead or deceive an elector in relation to the casting of a vote.**

There is a similar criminal provision in Section 122 of the *Referendum (Machinery Provisions) Act* 1984 around misleading or deceptive publications.

> **A person shall not, during the referendum period in relation to a referendum, print, publish or distribute, or cause, permit or authorise to be printed, published or distributed, any matter or thing that is likely to mislead or deceive an elector in relation to the casting of a vote at the referendum.**

There are otherwise no federal provisions nor explicit requirements for truth in political advertising. It is worth noting that Australia has experienced multiple iterations of public debate over 'truth in political advertising' legislation, with relevant state and territory legislation in place in South Australia (*Electoral Act* 1985 (SA), s 113) and the Australian Capital Territory (*Electoral Act* 1992 (ACT) s 297A). Typically, reform proposals centre on authorised election advertising by identifiable political actors. Our project suggests that *unauthorised* advertising stemming from beyond these candidate and partisan channels would also be worthy of consideration, should 'truth in political advertising' proposals aim to tackle online misinformation and disinformation.

**6.** Digi 2022 *Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2022/12/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-_-December-22-2022.docx.pdf

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

6

# The *Australian Code of Practice on Disinformation and Misinformation*

The Code is the key policy instrument underpinning multiple large platforms' responsibilities towards Australian users. Under the Code, platforms that sign on have obligations to develop and implement measures that aim to reduce the propagation and exposure of users to misinformation and disinformation. The Code includes two optional measures around:

**1. The monetisation of misinformation and disinformation. Platforms that have signed on to this optional objective are committing to:**

- using reasonable efforts to deter advertisers from repeatedly placing digital advertisements that propagate misinformation and disinformation; and

- implementing policies and processes that disrupt advertising and monetisation of misinformation and disinformation, which may include, for example:

  - developing and promoting brand safety and verification tools;

  - using third-party verification companies;

  - assisting advertisers in assessing media-buying strategies and online reputational risks;

  - providing advertisers with necessary access to client-specific accounts to enable them to monitor the placement of advertisements and make choices regarding where advertisements are placed;

  - restricting the availability of advertising services and paid placements on accounts and websites that propagate disinformation or misinformation.

**2. Improving public awareness of the source of political advertising carried on digital platforms. Platforms that have signed on to this optional objective are committing to:**

- Developing and implementing policies that provide users with greater transparency about the source of political advertising carried on digital platforms, which may include:

  - requirements that advertisers identify and/or verify the source of political advertising carried on digital platforms;

  - policies that prohibit advertising that misrepresents, deceives or conceals material information about the advertiser or the origin of the advertisement;

  - the provision of tools that enable users to understand whether a political ad has been targeted to them; and

  - policies requiring that political advertisements appearing in a medium containing news or editorial content are presented to be readily recognisable as a paid-for communication.

- Platforms may also choose not to target advertisements based on the inferred political affiliations of a user or develop other policies or processes around, for example, advertising that advocates for a political outcome on social issues of public concern.

Signatories to these optional measures include TikTok, Twitter (now X) and Facebook. Platforms are required to self-report against compliance with this commitment annually.

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

7

Political advertising may be excluded from the scope of the Code when it comes to misinformation. That is, the Code excludes all political advertising from being considered as misinformation. However, political advertising may fall under the definition of disinformation if it is propagated by manipulative, bulk aggressive behaviour, such as bots and spamming. This creates a grey area of protection. Political misinformation deliberately propagated through paid-for advertising created by individual actors may not be covered by the Code; only disinformation propagated by automated or bulk means is.

More details are available in Appendix 1.

### Under the *Australian Code of Practice on Disinformation and Misinformation,*

### Misinformation means:

1.  "Digital content (often legal) that is verifiably false or misleading or deceptive;

2.  Is propagated by users of digital platforms; and

3.  The dissemination of which is reasonably likely to (but may not be clearly intended to) lead to harm."

Political advertising is excluded from the definition of misinformation.

### Disinformation means:

1.  "Digital content that is verifiably false or misleading or deceptive;

2.  Is propagated amongst users of digital platforms via Inauthentic behaviours; and

3.  The dissemination of which is reasonably likely to cause harm."

Political advertising may be considered disinformation, if it is propagated by inauthentic behaviours.

Inauthentic behaviour "includes spam and other forms of deceptive, manipulative or bulk, aggressive behaviours (which may be perpetrated by automated systems) and includes behaviours which are intended to influence users' online conversations artificially and/or to encourage users of digital platforms to propagate Digital Content."

### Political advertising means paid advertisements:

1.  "Made by, on behalf of a political party, or

2.  That advocate for the outcome of an election, referendum or other Federal, State or Territory wide political process (such as a postal vote) supervised or managed by an electoral management body of the Commonwealth or State and Territory

3.  Are regulated as political advertising under Australian law."

### Harm is defined as harm that poses a credible and serious threat to:

1.  "Democratic, political and policy-making processes such as voter fraud, voter interference or voting misinformation or

2.  Public goods such as the protection of citizen's health, protection of marginalised or vulnerable groups, public safety and security of the environment."

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

8

# TikTok's policy and process

TikTok has relevant policies that prohibit paid-for political ads and those that should see the removal of electoral misinformation and disinformation post-fact-checking.

- TikTok does not allow paid political ads. It states that the platform does "not allow paid political promotion, political advertising, or fundraising by politicians and political parties (for themselves or others)".[7] This covers both 'traditional ads' paid for directly to the platform and 'organic ads', where creators post political advertising where they are paid for or compensated by government, politician or political party accounts (GPPPAs). However, the focus is on the actors who pay for the ads; that is, it prevents recognised GPPPAs from posting paid advertising. It does not prohibit paid-for ads that are political in nature from being posted by bad actors nor from propagating misinformation.

- TikTok removes electoral misinformation and disinformation, stating that it explicitly does not allow:

  - Inaccurate, misleading or false content that may cause significant harm to individuals or society, regardless of intent. Harm includes that done to democratic elections and their processes. Fact-checking is required to determine the inaccuracy or falseness of claims. Content that involves general conspiracy theories is allowed but not eligible for promotion in the For You Feed (the personalised feed targeted to individual users). TikTok allows statements of personal opinions as long as it does not include harmful misinformation.

  - Misinformation about the civil and electoral processes, regardless of intent. This includes inaccurate, misleading or false claims about how, when or where to vote; eligibility of voters; laws or procedures around votes; etc.

  - Paid political promotions, political advertising or fundraising by politicians or parties. [8]

This should cover electoral misinformation as propagated through paid-for advertising by non-GPPPAs.

TikTok states that 'advertisers on TikTok are responsible for their ads',[9] including complying with all relevant legislation and regulations and conforming with TikTok's policies. TikTok reviews all ads and implies that the review system is done by humans in its Ad Review FAQs (see Figure 1).[10] This could be intended to mean a 'human in the loop' automated system or complete human review.

## Does a person review my ads?

Yes, all ads are reviewed by us to ensure your ads follow the TikTok Community Guidelines and the TikTok Advertising Policies.

Figure 1: TikTok's FAQ about ad reviews, suggesting a 'person' reviews the ads

TikTok explains that "once ads are set up, ads will be reviewed automatically. The process will take about 24 hours. You can also trigger an ad review whenever you edit your ad group's targeting location or your ad's creative (images, ad text, links, videos, etc.)".

**See Appendix 3 for more details about TikTok's policies.**

---

7. TikTok 2023 *Civic and election integrity* https://www.tiktok.com/community-guidelines/en/integrity-authenticity

8. TikTok 2023 *Civic and election integrity* https://www.tiktok.com/community-guidelines/en/integrity-authenticity

9. TikTok 2023 *Advertising on TikTok - First Things to Note* https://ads.tiktok.com/help/article/advertising-on-tiktok-first-things-to-note?lang=en

10. TikTok 2023 *Ad review FAQs* https://ads.tiktok.com/help/article/ad-review-faq?lang=en

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

9

# Facebook's policy and process

Facebook has a range of overlapping relevant policies regarding political advertising and the propagation of misinformation and disinformation (see figure 2). Facebook states that the aim of these policies is "to help further protect people from poor experiences' and that they 'don't want ads that use profanity, show excessive nudity or include misinformation".

- Meta allows political advertising on its platforms, including Facebook. [11] It states, "Advertisers can run ads about social issues, elections or politics, provided that the advertiser complies with all applicable laws and the authorisation process required by Meta". Its definition of political advertising is neutral on who is paying for the ad and covers all advertising regarding elections, referendums or ballot initiatives. It states that "any advertiser running ads about social issues, elections or politics who is located in or targeting people in designated countries must complete (a Meta Authorisation process)'" and that "if ads do not include a disclaimer and we determine that the ad content includes content about social issues, elections or politics, it will be disapproved during ad review". This should prevent most electoral misinformation and disinformation being propagated in paid-for ads from bad actors.

- Meta's community guidelines [12] suggest that Facebook removes misinformation and disinformation in the following cases:

  - It is likely to contribute to the risk of imminent physical harm, including the risk of violence to people and harmful health misinformation, such as vaccine misinformation or the promotion of miracle cures.

  - It is highly deceptive media, such as deepfakes.

  - It is likely to directly contribute to interference with the functioning of political processes, including misinformation about the dates, locations, times and methods for voting, misinformation about who can vote, etc.

---

**11.** Meta 2023 *Ads about social issues, elections or politics* https://transparency.fb.com/en-gb/policies/ad-standards/siep-advertising/siep/

**12.** Meta 2023 *Community Standards: Misinformation* https://transparency.fb.com/en-gb/policies/community-standards/misinformation/

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

10

**EXAMPLES OF OTHER POLICIES THAT MAY APPLY**

**Branded Content Policies**
Ads may include branded content, which are subject to our Branded Content Policies.

**Developer Policies for Messenger**
Advertisers sending Sponsored Messages via Messenger Platform or starting automated conversations from ads that click to Messenger must comply with our Developer Policies.

**Pages, Groups and Events Policies**
Advertisers may promote Pages or Events, which are subject to the applicable Pages, Groups and Events Policies.

**Commerce Policies**
Advertisers may run certain types of ads, such as ads with product tags or boosted Marketplace listings, or use Catalogs to upload products for use in Commerce surfaces, which are subject to our Commerce Policies.

**Advertisers running Facebook ads are required to follow certain sets of policies**

**Advertising Policies**
All ads must comply

**Facebook Community Standards and Instagram Community Guidelines**
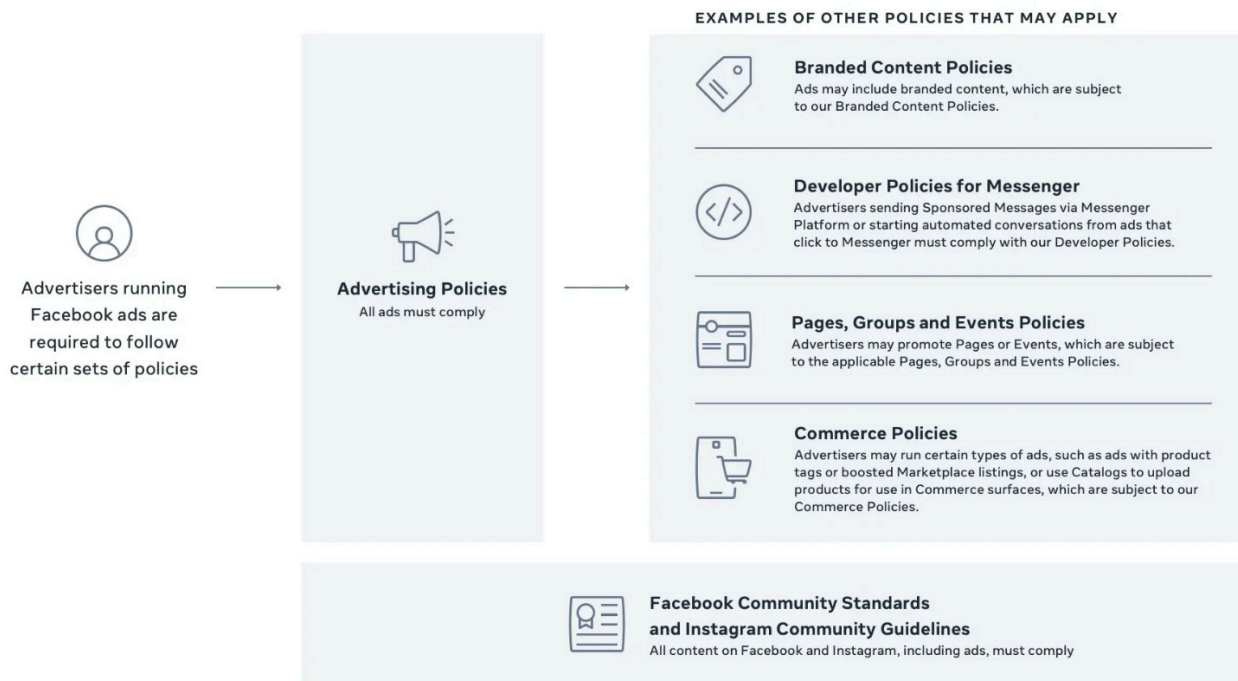All content on Facebook and Instagram, including ads, must comply

Figure 2: Facebook's policies regarding advertising. [13]

Facebook uses automated tools and systems to ensure compliance with its advertising policies. It describes its ad review system as follows:

"Our ad review system is designed to review all ads before they go live. This system relies primarily on automated technology to apply our Advertising Policies to the millions of ads that run across our apps. While our review is largely automated, we rely on our teams to build and train these systems, and in some cases, to manually review ads.… While ad review is typically completed within 24 hours, it may take longer, and ads can be reviewed again, including after they're live. Based on the results of the review, an ad is either rejected or allowed to run".[14]

We tested this system up until the 'allowed to run' stage (see Figure 3).

---

**13.** Meta 2021 *Breaking Down Facebook's Ad Review Process* https://www.facebook.com/business/news/facebook-ad-policy-process-and-review

**14.** Meta 2021 *Breaking Down Facebook's Ad Review Process* https://www.facebook.com/business/news/facebook-ad-policy-process-and-review

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
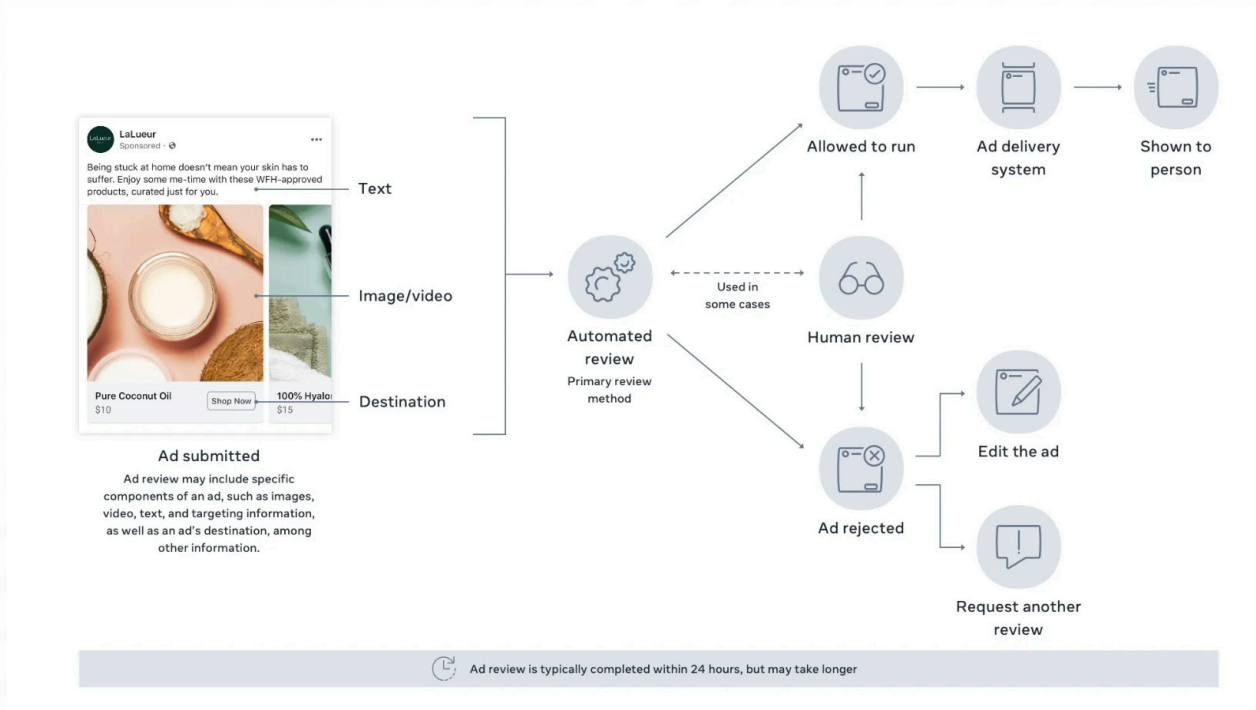**An experimental evaluation using the Voice referendum**    *11*

Figure 3: Facebook's policies regarding advertising.[15]

**See Appendix 3 for more details about Meta's policies.**

# X (Twitter)'s policy and process

X has relevant policies that should see the removal or labelling of electoral process misinformation and disinformation.

- X allows paid political content advertising,[16] although not in Australia. It states "Advertisers may not promote false or misleading content. This includes:

  - False or misleading information about how to participate in an election.

  - False or misleading information intended to intimidate or dissuade people from participating in an election.

  - False or misleading information intended to undermine public confidence in an election.

  Advertisers must comply with any applicable laws regarding disclosure and content requirements. Such compliance is the sole responsibility of the advertiser."

- X labels or removes electoral misinformation and disinformation,[17] stating: "You may not advance verifiably false or misleading information about how to participate in an election or other civic process. This includes but is not limited to... misleading information about procedures to participate in a civic process (for example, that you can vote by Post, text message, email, or phone call in jurisdictions where these are not a possibility)."

---

**15.** Meta 2021 *Breaking Down Facebook's Ad Review Process* https://www.facebook.com/business/news/facebook-ad-policy-process-and-review

**16.** X 2023 X *Ads Policies: Political Content* https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html

**17.** X 2023 *Civic integrity and misleading information policy* https://help.twitter.com/en/rules-and-policies/election-integrity-policy

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

12

X states, "Advertisers on X are responsible for their X Ads. This means following all applicable laws and regulations, creating honest ads, and advertising safely and respectfully." It uses automated tools to check that ads comply with its policies.

It states that "when advertisers on X choose to promote their content with X Ads, their account and content become subject to an approval process. The approval process is designed to support the quality and safety of the X Ads platform. This process helps X check that advertisers are complying with our advertising policies."

X affirms that ads "can be reviewed prior to running in campaigns. They are submitted for approval on an automatic basis, based on an account's advertising status, its historical use of X, and other evolving factors. Review generally takes into consideration how an account uses X, its profile, its content, and targeting included in any active or draft advertising campaigns."

Specifically, for promoted posts, X states that "posts can be marked for review once an advertiser selects them to be included in a Promoted Ads campaign. ... Once a post is approved for advertising, it immediately resumes running and can be included in new campaigns. When a post is disapproved for advertising, it cannot be included in X Ads campaigns.... Nothing about a post outside of X Ads changes because it is in review for advertising. Followers will see the post regardless of whether it is approved to be promoted. If an advertising account is found to be in violation of our Terms of Service, it may be suspended from the service in addition to being removed from the X Ads platform."

**See Appendix 4 for more details about X's policies.**

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**    13

# Method

While it is unclear from the policy landscape what the response to electoral misinformation and disinformation propagated in paid-for ads should be, we set out to see what the response was in reality. We did this by conducting a simple, small experiment following four steps:

**1. We developed a set of electoral process misinformation paid-for ads, including content suggesting that:**

- the date of the Voice referendum was set to 31 November, both incorrect and an obviously non-existent date;

- the Voice referendum was voluntary and/or a postal survey;

- it was possible to take part in the Voice referendum by SMS;

- 16-year-olds could vote;

- the Voice referendum would use electronic voting systems;

- the Voice referendum had to be cancelled; and

- the AEC had declared support for either Yes or No campaigns and/or that they had been found guilty of electoral fraud.

**2. We established Facebook, TikTok and X (Twitter) accounts, which were clearly labelled as being run by Reset.Tech Australia or staff.**

- On TikTok, we established a non-GPPPA.

- On Facebook, we did not apply for authorisation to run political ads.

- On X, the account was 'Blue Tick' verified and based in Europe with an Australian focus and audience.

**3. We put forward these electoral process misinformation paid-for ads for platform approval to see if platforms would approve them.**

**4. We then cancelled our application for these ads and posts to ensure they did not run.** To be clear, we did not run any of these ads, and no one was exposed to electoral process misinformation as part of this experiment. These ads were only visible to Reset.Tech researchers and the platforms' automated checking systems.

There were limitations to this method. Because we cancelled these ads, we could not be sure if they would have been caught later in the approval process or how quickly they may have been reported and taken down. All we could stress test was the adequacy of the ad-approval process described and delivered by platforms. Previous experimental research conducted by Reset.Tech Australia has suggested no further steps in moderation between the 'ad approval' process and the ads running on Facebook.[18]

---

**18.** Dylan Williams 2020 *How Facebook lets you break Australian electoral laws in under 15 minutes* https://medium.com/ausreset/how-facebook-lets-you-break-australian-electoral-laws-in-under-15-minutes-7db5619ccc9b

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

14

# We have done this before

Repetition is a key part of stress testing platform policies and processes and vital to documenting improvements or regressions. We have repeated these methods twice before on Meta platforms:

- In 2020, we tested if Meta would approve and run electoral misinformation in paid-for ads. In this instance, we recruited a hundred people for the experiment, who agreed to receive these ads. Using Facebook's targeting techniques, we were able to exclusively target these hundred people alone, who reported that they were indeed delivered.[19]

- In the run-up to the 2022 federal election, we found that five out of five election misinformation ads that we had placed on Meta were approved to run and scheduled for delivery. Four of them were approved in under three hours, and one was approved in less than twenty-four hours.[20]

This research set out to see if these processes had improved on Facebook and to also stress test TikTok and X's processes.

---

**19.** Dylan Williams 2020 How Facebook lets you break Australian electoral laws in under 15 minutes https://medium.com/ausreset/how-facebook-lets-you-break-australian-electoral-laws-in-under-15-minutes-7db5619ccc9b

**20.** Dylan Williams & Dhakshayini Sooriyakumaran 2022 *Facebook still approving ads with explicit disinformation as election campaigning ramps up* https://au.reset.tech/uploads/facebook-electoral-disinfo-ad-experiment-1.pdf

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

15

# Findings

## Tiktok

We put forward 10 paid-for ads containing electoral misinformation for TikTok to review. TikTok's ad approval system requires advertisers to upload and submit their ads for review. We made ten ads for TikTok, all consisting of images of electoral misinformation and political content regarding the Voice referendum. We used a TikTok tool to cluster these images into a video ad. TikTok then clustered the ads into sets of three to run.

Seven of these ads containing a range of misinformation were approved by TikTok's system in under twenty-four hours (see Figure 5). One ad was detected and rejected by TikTok, who notified us by stating the following:

> **"Rejection reason for ad [number] (show up to 3 ads): The ad or video features political content. This could include, but is not limited to: - Ads that reference, promote, or oppose candidates or nominees for public office, political parties, or elected or appointed government officials; - Ads that reference an election, including voter registration, voter turnout, and appeals for votes; - Ads that include advocacy for or against past, current, or proposed referenda, ballot measures, and legislative, judicial or regulatory outcomes or processes; - Ads that reference, promote, sell, merchandise that feature prohibited individuals, entities or content, including campaign slogans, symbols, or logos."**

Because this ad image was rejected and grouped in a batch of three images, the final two ad images were not reviewed by TikTok and have neither been approved nor rejected.

Our account was listed as "needs attention" as a result of this ad being disapproved but was still active and able to run ads, and we were, in principle, still able to run the other seven ads. TikTok's automated system sent the following message:

> **"We noticed that one or more ad groups under the ad account [number] have violations of TikTok Advertising Policies. It has caused the overall ad account health status updated from "Good" to "Attention needed."**
>
> **Ad accounts may face potential suspension and lose access to ad campaigns, followers, and funds when your ad account health status turns "Poor."**

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

16

An ad claiming that this 'weird old tip' could get you two votes in the Voice


An ad claiming that voting was available via SMS


An ad claiming that voting was available via SMS


An ad claiming that the date for the Voice was November 31


An ad claiming that the vote is rigged and discouraging participation


An ad claiming that it is a scam and encouraging informal voting


An ad claiming that the referendum was optional

Figure 4: The 7 ads that were all approved by TikTok's ad approval system



Figure 5: The ad that was disapproved by TikTok's ad approval system

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

17

While political advertising is explicitly excluded from the Code, TikTok has committed to the two relevant optional commitments. In TikTok's first annual transparency report regarding its compliance, it states the following:

1.  Regarding disruption of advertising for and monetising misinformation and disinformation, it states "that the platform has strong platform controls and does not allow political advertising" (see Figure 6).[21]

2.  Regarding improving public awareness of the source of political advertising carried on digital platforms, "TikTok strictly prohibits political and issue-based advertising, as per TikTok advertising policies…. Disruption of monetisation incentives for political content and GPPPA verification helps TikTok enable political discussions in a safe, entertaining environment."

Nevertheless, this does not align with this research's findings. We were able to approve in principle that paid-for ads were political in nature and included electoral misinformation. We received an account warning, but the account remained active and could advertise.

### Disrupt advertising and monetisation incentives for disinformation.

**"Outcome 2: Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.**

As TikTok grows, we continue to maintain strong platform control by strengthening our advertising policies. We do not allow the monetisation of government-owned accounts or political advertising, with the exception of cause-based advertising and information notices from non-profit or governmental organisations in collaboration with TikTok Sales Representatives. Our advertising policies also contain strict prohibitions on ads that contain deceptive or misleading claims, or which attempt to exploit or profiteer from sensitive events or subjects, such as the COVID-19 pandemic."

Figure 6: An excerpt from TikTok's annual transparency report to Digi regarding advertising and disinformation

---

21. TikTok 2023 *Annual Transparency Report TikTok Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2023/05/TikTok-2022-Annual-Transparency-Report.pdf

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

*18*

# Facebook

We made twenty ads for Facebook to review, all consisting of images of electoral misinformation and political content regarding the Voice referendum. Facebook's ad approval system asks advertisers to self-identify ads that are about social issues, elections or politics. We tested one ad using this category, which Meta correctly rejected six hours later because we had not gone through Meta's account authorisation system to enable us to run ads about social issues, elections or politics.

We tested a further nineteen paid-for ads without self-identifying them as ads that are about social issues, elections or politics. All nineteen ads were approved within six to twenty-four hours. These ads contained a variety of electoral misinformation (See Figure 7).

Our account received no warnings and was still active at the time of publication.



An ad claiming that 16-year-olds can vote in the referendum



An ad claiming that voting was available via SMS



An ad claiming that voting was available via SMS



An ad claiming that Indigenous and Torres Straight (sic) Islanders can vote twice



An ad claiming that the date for the Voice was November 31



An ad claiming that electronic voter machines would be used in the Voice referendum



An ad claiming that the Voice is postal and voluntary



An ad discouraging voting, saying your vote will not be counted



An ad claiming that the Voice is optional

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

*19*

An ad claiming that the AEC was supporting the Yes campaign


An ad claiming that the AEC was supporting the No campaign


An ad claiming that the election would not be overseen by the AEC


An ad claiming that the AEC was under investigation for fraud


An ad claiming that the AEC was under investigation for fraud
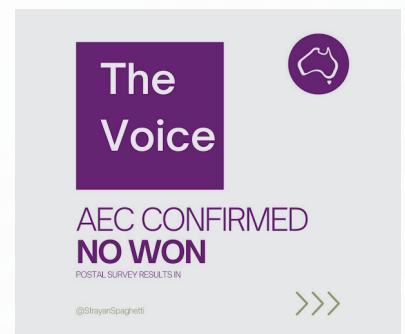

An ad claiming urging people not to take part in the referendum because Australia is governed by maritime law


An ad claiming that the referendum results had already been counted and that Yes won


An ad claiming that the referendum results had already been counted and that Yes won


An ad claiming that the referendum results had already been counted and that No won


An ad claiming that the referendum had been cancelled

Figure 7: The nineteen ads that were all approved by Facebook's ad approval system

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

20

While political advertising is explicitly excluded from Digi's Code, Meta has committed to the two relevant optional commitments. In Meta's first annual transparency report regarding its compliance, it states the following:
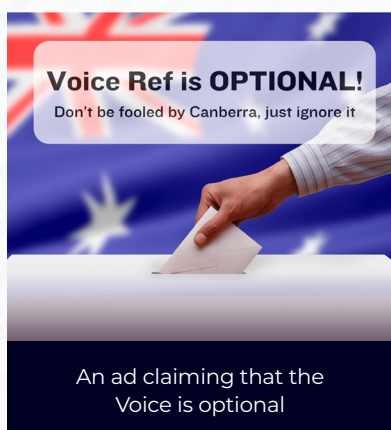
1. Regarding disruption of advertising for and the monetisation of misinformation and disinformation, "Meta sets a higher threshold for users to be able to advertise on our services, and takes action against users who spread misinformation."[22]

2. Regarding improving public awareness of the source of political advertising carried on digital platforms, "Meta requires all advertisers of political and social issue ads to complete an ad authorisation, which includes verifying the advertiser's identity. Meta requires political and social issue ads to include a disclaimer disclosing who is paying for the ad. Meta provides the Ad Library, a searchable archive of all political and social issue ads on our services in Australia. Meta enables an Ad Library report that provides aggregated spend information about Pages undertaking political and social issue ads."

Nevertheless, this does not align with this research's findings. We were able to approve in principle that paid-for ads were political – without completing an ad authorisation – and included electoral misinformation.

# X (Twitter)

We made fifteen ads for X, all consisting of images of electoral misinformation and political content regarding the Voice referendum. X's ad approval system – for personal accounts rather than business accounts – allows users to submit 'posts' to be boosted as paid-for advertising in users' feeds. We made fifteen different posts for X, all of which consisted of images of electoral misinformation and political content regarding the Voice referendum (see Figure 8). All of these were submitted for approval and scheduled to run.

Our account received no warnings and was still active at the time of publication.



An ad claiming that the Voice is optional



An ad claiming that the date for the Voice was November 31



An ad claiming that the referendum was a postal survey with no clear date

---

22. TikTok 2023 *Annual Transparency Report TikTok Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2023/05/TikTok-2022-Annual-Transparency-Report.pdf
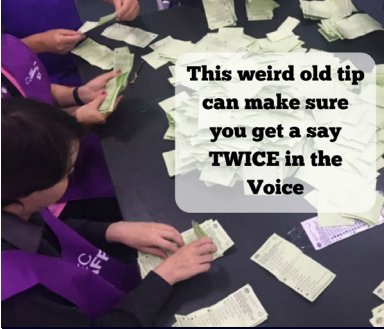
Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

21

An ad claiming that voting was available via SMS


An ad claiming that 'this weird old tip' will give you two votes in the Voice


An ad claiming that you can 'skirt' the AEC and vote twice


An ad claiming that the referendum had been cancelled


An ad claiming that the AEC was under investigation for fraud


An ad claiming that 16-year-olds can vote in the referendum


An ad urging people not to take part in the referendum because Australia is governed by maritime law


An ad claiming that the Voice result was already decided, and that voting doesn't matter


An ad claiming that the AEC won't be overseeing the referendum


An ad claiming that the AEC was supporting the Yes campaign


An ad claiming that the AEC was supporting the Yes campaign


An ad claiming that the referendum was voluntary and by post

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
**An experimental evaluation using the Voice referendum**

22

While political advertising is explicitly excluded from the Code, X has committed to the two relevant optional commitments. In X's first annual transparency report regarding its compliance, it states the following:

1. Regarding disruption of advertising for and the monetisation of misinformation and disinformation, "promoted content must adhere to their rules and policies" (see Figure 9).[23]

2. Regarding improving public awareness of the source of political advertising carried on digital platforms, it is "not applicable in Australia", presumably because X does not allow promoted political content in Australia in its rules.

However, this is not in line with this research's findings. We were able to approve in principle that paid-for posts included relevant Australian political content and electoral misinformation.

## Disrupt advertising and monetisation incentives for disinformation.

"As reported above, as we evolve, we are giving people greater transparency and control over their experience on the platform, and this includes our advertisers. We have also been working to improve the advertising experience on Twitter by making ads more relevant. Underpinning these efforts is our work to ensure ads appear in brand-suitable environments. Ensuring that the context in which ads appear does not conflict with a brand's message and values is foundational to delivering a safe, relevant, and informative experience for everyone on Twitter.

Promoted content on Twitter must adhere to the Twitter rules and our advertising policies. People using Twitter can also make reports related to Twitter Ads that might potentially violate our policies. These will be assessed against the Twitter Ads Policy, the Twitter Rules and TOS and any enforcement action will be taken in line with these policies. As mentioned above, Twitter uses a combination of human review and technology to help us enforce our rules. Our specially trained team reviews and responds to reports 24/7; they have the capacity to review within context and respond to reports in multiple languages. In addition, we publish specific policies for advertisers that share standards for (sic) that are outlined below.

- **Political content advertising policy:** During the reporting period we continued to prohibit political advertising in Australia and as reflected under our political content advertising policy. We'll continue to review and update these.

- **Inappropriate content advertising policy:** Our policy on inappropriate content advertising prohibits advertising deemed to be dangerous or exploitative, misrepresentative, along with misleading synthetic or manipulated content and content engaged in coordinated harmful activity.

- **Quality advertising policy:** Our quality advertising policy outlines standards for advertisers including that ads should represent the brand or product being promoted and cannot mislead users into opening content by including exaggerated or sensationalised language or misleading calls to action."

Figure 9: An excerpt from X's (then Twitter) annual transparency report to Digi regarding advertising and disinformation

---

**23.** Twitter 2023 *Australian Code of Practice on Disinformation and Misinformation* Twitter Annual Transparency Report Reporting Period: 2022 https://digi.org.au/wp-content/uploads/2023/05/ACPDM_report_2022_Twitter-052823_DIGI.pdf

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

23

# Conclusion

- This research demonstrated the significant gaps in Australia's regulatory framework when it comes to electoral misinformation and disinformation served through paid-for advertising.

- The regulatory framework for electoral misinformation and disinformation is unclear, and platform policies often present loopholes. They fail to stop bad actors.

- Regardless, it is clear that, in practice, it is too easy to get electoral misinformation propagated in paid-for ads. The ad approval systems deployed by platforms are clearly not adequate and fail to detect electoral misinformation.

  - Facebook's system appeared to be dependent on advertiser self-declarations regarding the nature of the advertising, which offers insufficient protection against bad actors. We submitted twenty ads containing paid-for misinformation to test Meta's ad approval system, and 95% were approved. Meta approved all nineteen ads that were not self-identified as political ads, rejecting only one ad that we had voluntarily identified as a political ad.

  - TikTok's system appeared to catch some political advertising and misinformation, but not the majority. We submitted ten ads to test TikTok's ad approval system, and 70% were approved. TikTok approved seven ads, rejected one ad and did not review the final two after detecting the violating ad.

  - X's (Twitter's) system did not detect or reject electoral process misinformation in paid-for advertising. We submitted fifteen ads to test X's ad approval system, and 100% were approved and scheduled to run.

- None of the platform's transparency reports under the voluntary Code adequately identified nor addressed these issues. This small experiment suggests that self-reporting mechanisms under Digi's Code may be weak and require more effective scrutiny.

- There are clearly vulnerabilities in Australia's online information architecture, leaving us exposed to attacks from bad actors.

- On the policy side, there is a clear need for more concise regulation and stronger enforcement:

  - We do not recommend that electoral misinformation be included in the voluntary Code, given the overall ineffectiveness of the Code's implementation and enforcement.

  - Section 122 of the *Referendum (Machinery Provisions) Act* 1984 does not appear to be acting as an effective deterrent to prevent platforms publishing electoral process misinformation in paid-for advertising.

  - There is a need for a strong regulatory framework. The European experience, summarised in appendix 5, outlines how other jurisdictions identified the need to and successfully implemented the move from faulty voluntary Codes to effective regulations leading to the *Digital Services Act*. As the government considers the next steps regarding the *Exposure Draft Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill* and reviews the *Online Safety Act*, consideration needs to be given to a more comprehensive risk-based approach to regulations.

- On the platform's side, there is a clear need for more rigorous policies and effective implementation.

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*    24

# Appendices

## Appendix 1: Digi's Australian Code of Practice on Disinformation and Misinformation

There is an optional objective for signatories under Digi's Code that is relevant to monetisation and another around improving public awareness of political advertising carried on platforms. These are summarised below by Reset.Tech. Other compulsory and optional commitments are available on Digi's website for review.[24]

**Objective: Disrupt advertising and monetisation incentives for Disinformation and Misinformation.**

**Outcome: Advertising and/or monetisation incentives for Disinformation and Misinformation are reduced.**

- Signatories that offer digital advertising services will use commercially reasonable efforts to deter advertisers from repeatedly placing digital advertisements that propagate Disinformation or Misinformation.

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

- Policies and processes implemented (to achieve this) may for example, include:

  - promotion and/or inclusion of the use of brand safety and verification tools;

  - enabling engagement with third party verification companies;

  - assisting and/or allowing advertisers to assess media buying strategies and online reputational risks;

  - providing advertisers with necessary access to client-specific accounts to help enable them to monitor the placement of advertisements and make choices regarding where advertisements are placed; and /or

  - restricting the availability of advertising services and paid placements on accounts and websites that propagate Disinformation or Misinformation.

- Signatories recognise that all parties involved in the buying and selling of online advertising and the provision of advertising-related services need to work together to improve transparency across the online advertising ecosystem and thereby to effectively scrutinise, control and limit the placement of advertising on accounts and websites that propagate Disinformation

---

**24.** Digi 2022 *Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2022/12/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-_-December-22-2022.docx.pdf

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid–for advertising?*
*An experimental evaluation using the Voice referendum*

25

**Objective: Improve public awareness of the source of Political Advertising carried on digital platforms.**

**Outcome Users are better informed about the source of Political Advertising.**

- While Political Advertising is not Misinformation for the purposes of the Code, Signatories will develop and implement policies that provide users with greater transparency about the source of Political Advertising carried on digital platforms.

- Measures developed and implemented in accordance with the commitment may include requirements that advertisers identify and/or verify the source of Political Advertising carried on digital platforms; policies which prohibit advertising that misrepresents, deceives, or conceals material information about the advertiser or the origin of the advertisement; the provision of tools which enable users to understand whether a political ad has been targeted to them; and policies which require that Political Advertisements which appear in a medium containing news or editorial content are presented in such a way as to be readily recognisable as a paid-for communication.

- Signatories may also, as a matter of policy, choose not to target advertisements based on the inferred political affiliations of a user or choose to define and implement commitments concerning a broader scope of political advertising including advertising that advocates for a political outcome on social issues of public concern.

# Appendix 2: TikTok's policies

TikTok has policies that restrict the actions of GPPPAs on platforms on parties and aim to limit misinformation and disinformation on its platforms. Relevant summaries of these are provided below.

## 1. Restrictions on Governments, politicians and political parties[25]

TikTok's description of a GPPPA includes:
1. national/federal government-run entities, such as agencies/ministries/offices;
2. state/provincial and local government entities;
3. candidates and elected officials at the federal/national level;
4. government officials at the federal/national level, such as cabinet ministers and ambassadors;
5. official spokesperson or member of senior staff for a national/state level candidate or elected/appointed official, such as the chief of staff, campaign director or digital director;
6. official spokesperson, member of senior staff or executive leader for a political party, such as the party chairman or finance director;
7. political parties;
8. royal family members with official government capacities;
9. political youth associations (for main political parties at the discretion of regional public policy);
10. former heads of state or heads of government;
11. political action committees (PACs) or any country-specific equivalents;
12. candidates and elected officials at the state/provincial and local levels as determined by regional public policy based on market factors; and
13. government officials at the state/provincial and local levels as determined by regional public policy based on market factors.

---

**25.** TikTok 2023 *Government, Politician, and Political Party Accounts* https://support.tiktok.com/en/using-tiktok/growing-your-audience/government-politician-and-political-party-accounts

Reset. AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising? An experimental evaluation using the Voice referendum*

26

TikTok prohibits GPPPA accounts from advertising. "TikTok has long prohibited political advertising, including both paid ads and creators being paid to make branded political content. This also includes the use of promotional tools available on the platform, like Promote or TikTok Shop. In addition to our political advertising content policy, we also impose prohibitions at the account level. This means that accounts we identify as belonging to politicians and political parties have their access to advertising features turned off. We recognize that there will be occasions when official government organizations may have reason to advertise on TikTok, for example, to promote public health initiatives like COVID-19 booster programs. We will continue to allow government organizations to advertise in limited circumstances, and they are required to work with a TikTok representative."

## 2. Restrictions on propagating misinformation and disinformation[26]

TikTok's community guidelines state that it removes violative content from the platform that breaks its rules. This includes misinformation content that can cause significant harm, including harm that affects electoral processes.

"We do not allow inaccurate, misleading, or false content that may cause significant harm to individuals or society, regardless of intent. Significant harm includes physical, psychological, or societal harm, and property damage."

"We do not allow misinformation about civic and electoral processes, regardless of intent. This includes misinformation about how to vote, registering to vote, eligibility requirements of candidates, the processes to count ballots and certify elections, and the final outcome of an election. Content is ineligible for the FYF if it contains unverified claims about the outcome of an election."

# Appendix 3: Meta's policies

Meta has policies regarding ads about social issues, elections or politics and others that aim to restrict misinformation and disinformation from its platforms. Relevant summaries of these are provided below.

## 1. Ads regarding social issues, elections or politics[27]

Meta allows political advertising on its platforms, including Facebook. It states "advertisers can run ads about social issues, elections or politics, provided that the advertiser complies with all applicable laws and the authorisation process required by Meta. Meta may restrict issue, electoral or political ads. In addition, certain content related to elections may be prohibited by local law or removed in specific regions ahead of voting."

---

**26.** TikTok 2023 *Civic and election integrity* https://www.tiktok.com/community-guidelines/en/integrity-authenticity/s

**27.** Meta 2023 *Ads about social issues, elections or politics* https://transparency.fb.com/en-gb/policies/ad-standards/siep-advertising/siep/

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

27

The policy states that "any advertiser running ads about social issues, elections or politics who is located in or targeting people in designated countries must complete (a Meta Authorisation process), except for news publishers identified by Meta." This applies to any ad that:

- "Is 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

- Is about any election, referendum or ballot initiative, including "get out the vote" or election information campaigns

- Is about any social issue in any place where the ad is being run

- is regulated as political advertising."

"Advertisers must include a verified "Paid for by" disclaimer on these ads to show the entity or person responsible for running the ad across Meta technologies... Advertisers running these ads, regardless of location, targeting or news exemptions, must comply with all applicable laws and regulations...

If ads do not include a disclaimer and we determine that the ad content includes content about social issues, elections or politics, it will be disapproved during ad review. If an ad is already running, it can be flagged by automated systems or reported by our community and, if found to be violating our policy by missing a disclaimer, it will be disapproved and added to the Ad Library."

## 2. Restrictions on propagating Misinformation and Disinformation[28]

Meta's community guidelines suggest that Facebook removes mis and dis information where:

- It is likely to contribute to the risk of imminent physical harm, including risk of violence to people, harmful health mis information including vaccine misinformation or the promotion of miracle cures for example

- It is highly deceptive media, such as deepfakes, or

- It is likely to directly contribute to interference with the functioning of political processes, as detailed below.

---

28. Meta 2023 *Community Standards: Misinformation* https://transparency.fb.com/en-gb/policies/community-standards/misinformation/

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*          28

"In an effort to promote election and census integrity, we remove misinformation that is likely to directly contribute to a risk of interference with people's ability to participate in those processes. This includes the following:

- Misinformation about the dates, locations, times and methods for voting, voter registration or census participation.

- Misinformation about who can vote, qualifications for voting, whether a vote will be counted and what information or materials must be provided in order to vote.

- Misinformation about whether a candidate is running or not.

- Misinformation about who can participate in the census and what information or materials must be provided in order to participate.

- Misinformation about government involvement in the census, including, where applicable, that an individual's census information will be shared with another (non-census) government agency.

- Content falsely claiming that the US Immigration and Customs Enforcement (ICE) is at a voting location.

- Explicit false claims that people will be infected by COVID-19 (or another communicable disease) if they participate in the voting process."

However, they go on to state that "For all other misinformation, we focus on reducing its prevalence or creating an environment that fosters a productive dialogue."

# Appendix 4: X's (Twitter's) policies

X has policies to restrict misinformation and disinformation from its platforms and regarding political advertising. Relevant summaries of these are provided below.

## 1. Restrictions on political advertising[29]

X permits the promotion of political advertising targeting specified countries with restrictions. It states: "Political ads should comply with the country specific legal requirements, election laws and applicable electoral silence periods mandates".

**What counts as political advertising?**

"Political Content ads and Political Campaigning ads are subject to this policy:

- Political content ads are defined as ads that reference a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome

---

**29.** X 2023 *X Ads Policies: Political Content* https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

29

- Political campaigning ads are defined as:
  - Ads that advocate for or against a candidate or political party.
  - Ads that appeal directly for votes in an election, referendum, or ballot measure.
  - Ads that solicit financial support for an election, referendum, or ballot measure.
  - Ads from registered PACs and SuperPACs.

**How does this apply to Australia?**

"Political Content ads are permitted in the following countries: Argentina, Brazil, Canada, Chile, Ecuador, Israel, Japan, Mexico, Peru, United Kingdom and the United States... subject to following restrictions:

- Advertisers may not promote false or misleading content. This includes:
  - False or misleading information about how to participate in an election.
  - False or misleading information intended to intimidate or dissuade people from participating in an election.
  - False or misleading information intended to undermine public confidence in an election.
- Advertisers must comply with any applicable laws regarding disclosure and content requirements. Such compliance is the sole responsibility of the advertiser.

Political campaigning ads are permitted in the following countries: Argentina, Japan, Mexico, United Kingdom and the United States. Advertisers must obtain pre-approval to run political ads by first getting certified.

Advertisements are subject to additional eligibility, product, and targeting restrictions:

- Political campaigning ads targeted to a country can only be bought by citizens or entities based out of that country
- Advertisers must comply with the country specific election laws (including campaign and silent periods)
- Political campaigning ads may only be promoted via the following ad formats: Promoted Ads, Follower Ads, X Amplify, X Takeover, X Live, and X Ad features; no other X advertising products or units are permitted
- All advertisers must comply with Twitter's Sensitive Categories Targeting Policy. Only the following criteria may be used to target political campaigning ads:
  - Location
  - Age
  - Gender
  - Interests & Keyword
  - Custom Audiences
  - Follower Look-alikes

Note: Political campaigning ads are currently supported on desktop and iOS and will only serve on those devices. We anticipate enabling political campaigns ads on Android and other devices in the coming weeks."

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

*30*

## 2.  Restrictions on propagating misinformation and disinformation[30]

Meta has a policy of removing or labelling electoral process misinformation content.

"You may not use X's services for the purpose of manipulating or interfering in elections or other civic processes, such as posting or sharing content that may suppress participation, mislead people about when, where, or how to participate in a civic process, or lead to offline violence during an election. Any attempt to undermine the integrity of civic participation undermines our core tenets of freedom of expression and as a result, we will apply labels to violative posts informing users that the content is misleading.

Information about a civic process is explained below.

"X considers civic processes to be events or procedures mandated, organized, and conducted by the governing and/or electoral body of a country, state, region, district, or municipality to address a matter of common concern through public participation. Some examples of civic processes may include:

- Political elections
- Censuses
- Major referenda and ballot initiatives"

Misleading information about how to participate is explained below.

"You may not advance verifiably false or misleading information about how to participate in an election or other civic process. This includes but is not limited to:

- misleading information about procedures to participate in a civic process (for example, that you can vote by Post, text message, email, or phone call in jurisdictions where these are not a possibility);

- misleading information about requirements for participation, including identification or citizenship requirements;

- misleading claims that cause confusion about the established laws, regulations, procedures, and methods of a civic process, or about the actions of officials or entities executing those civic processes; and

- misleading statements or information about the official, announced date or time of a civic process."

---

**30.**  X 2023 *Civic integrity and misleading information policy* https://help.twitter.com/en/rules-and-policies/election-integrity-policy

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*  *31*

# Appendix 5: The move from voluntary codes on misinformation and disinformation to regulation in Europe

The European experience shows how legislators have gradually responded to the shortcomings of the voluntary industry codes with a more comprehensive package. Notably, requirements for data access were consistently invoked to ensure that there were mechanisms for independent assessments of what was otherwise mere platform self-reporting.

| March 2018 | April 2018 | September 2018 | January 2019 | March 2019 | April 2019 |
|---|---|---|---|---|---|
| A final report of the High Level Expert Group on fake news and online disinformation | European Commission responds with a '*Code of Practice on Disinformation*' that would commit online platforms and the advertising industry to provide academia with 'access to platform data'. | Version 1 of the Code of Practice is released. | The European Commission expresses concern about the platforms' failure to benchmark and meaningfully measure progress. | The European Commission remarks platforms 'didn't provide access to more granular data to assess the effectiveness of their activities to counter disinformation'. | The European Commission calls for independent data access to ensure that the platforms are 'not just marking their own homework'. |

| 2019–2020 | September 2020 | 2020–2021 | June 2022 | November 2022 | September 2023 |
|---|---|---|---|---|---|
| An independent assessment by the European Regulators Group for Audiovisual Media Services notes no sufficient progress was made on platform commitments under the Code. | Findings from the European Commission on the first twelve months of the Code of Practice were released, noting 'shortcomings mainly due to the Code's self-regulatory nature'. | Draft *Digital Services Act* provisions construct a data access regime with a legal basis to force VLOPs/VLOSE to provide access to data to third parties, including regulators, vetted researchers and civil society organisations. | Roll-out of the 'Strengthened' Code of Practice on Disinformation | The *Digital Services Act* enters into force, including risk mitigation duties on platforms and mandated data access for regulators, civil society organisations and accredited researchers. | The first risk mitigation reporting from platforms expected under the *Digital Services Act* |

Reset.
AUSTRALIA

*How do platforms handle electoral misinformation in paid-for advertising?*
*An experimental evaluation using the Voice referendum*

32