

Reset.
AUSTRALIA

How do platforms' recommender systems promote political content?

An experimental investigation using the Voice referendum

Reset.Tech Australia
October 2023



Summary

Digital platforms are shaping the landscape of Australian political discourse. While significant attention is rightly paid to how platforms influence political discourse through content moderation and policies around misinformation and disinformation, they also shape political discourse through the development and deployment of algorithms in content recommender systems.

Recommender systems can distort political debate by promoting extremist¹ or dangerous content,² but can also shape debate by pushing one-sided or partisan content to users. This is often described as the ‘filter bubble’ effect or as social media ‘rabbit holes’, which can damage the plurality of the content people consume.

This research explores the effect of social media algorithms on political content promotion concerning the Voice referendum in Australia. We set up sock puppets (or ‘fake accounts’) on TikTok and X (formerly Twitter) to observe the rate at which these accounts fell into ‘Yes’ or ‘No’ filter bubbles.

Our findings include the following:

- **On TikTok:**
We primed four sock puppet accounts. Two of them fell into strong ‘No’ filter bubbles within 400 videos. One fell into a ‘Yes’ filter bubble within 250 videos, and one failed to fall into a filter bubble.
- **On X (formerly Twitter):**
We primed two sock puppet accounts, with one falling into a ‘No’ filter bubble after around 300 Xs (tweets) and the other into a ‘Yes’ filter bubble after around 200 Xs.

The existence of ‘Yes’ and ‘No’ filter bubbles, which can rapidly appear, suggests that platforms’ recommender systems could play a role in dividing Australian political discourse. This division could shape the polarities of Australian political debates.

Despite this, algorithms and content recommender systems remain largely invisible to Australian researchers, as platforms’ ‘transparency tools’ (APIs) are being closed down, moved behind paywalls, or are only available to Europeans or Americans. As the Government considers the next steps regarding the *Exposure Draft Communications Legislation Amendment (Combating Misinformation and Disinformation) Bill* and reviews the *Online Safety Act 2021*, consideration must be given to ensuring that independent oversight of algorithms is possible for regulators and researchers.

1. Ralph Housego & Rys Farthing 2022 ‘Social Grooming’ *AQ Magazine* <https://www.jstor.org/stable/27161413>

2. Reset.Tech Australia & ISD 2022 *Algorithms as a weapon against women* <https://au.reset.tech/uploads/algorithms-as-a-weapon-against-women-reset-australia.pdf>



Reset.
AUSTRALIA

Reset.Tech Australia is an Australian policy development and research organisation. We specialise in independent and original research into the social impacts of tech companies. We are the Australian affiliate of Reset.Tech, a global initiative working to counter digital harms and threats. Reset.Tech has extensive, global experience in monitoring electoral misinformation and disinformation with a focus on identifying areas for regulatory intervention. We are not affiliated to either referendum campaign.

Table Of Contents

Introduction	4
Method	6
Findings	7
TikTok	7
X (formerly Twitter)	8
Conclusions	8

Dr Rys Farthing, Noah Vatou, Aruna Anderson and Racheline Tantular

Authors & Researchers

Dr Rys Farthing, Noah Vatou, Aruna Anderson and Racheline Tantular

Cover artwork

Jamillah Knowles/Reset.Tech Australia/ Better Images of AI/ Connected People Australia / CC-BY 4.0

Funding



Introduction

Digital platforms are shaping the landscape of Australian political discourse. While significant attention is rightly paid to how platforms influence political discourse through content moderation and policies around misinformation and disinformation, they also shape political discourse through the development and deployment of algorithms in content recommender systems.

Content recommender systems employ powerful algorithms to recommend content to users, including political content. Their influence on shaping political debate should not be underestimated. Although platforms are often tight-lipped about specifics, at one stage, YouTube executives revealed that their recommender system drives 70% of the media that users consume on the platform.³

Recommender systems can distort political debate by promoting extremist⁴ or dangerous content,⁵ but can also shape debate by pushing one-sided or partisan content to users. This is often described as the ‘filter bubble’ effect or as social media ‘rabbit holes’, which can damage the plurality

of the content people consume. At the extreme end, there are extensive academic debates about whether filter bubbles can affect users and fuel radicalism,⁶ but on a more day-to-day level, they can leave political discussions more divided and without common ground.

Despite their role in shaping political discourse, there is limited visibility and transparency regarding these algorithms and their impact. X recently closed its main free ‘access tool’ API, which allowed some oversight, and now charges for access, while TikTok only offers an API that is accessible to American and European researchers at present. Similarly, Facebook has announced it is closing down its research tool ‘CrowdTangle’.⁷ Some jurisdictions, such as the EU and the UK, have regulatory requirements for digital platforms to share this information with vetted researchers, but Australia does not mandate such transparency. Mandated data access in other jurisdictions does not allow for research on harms to Australian users, leaving a blind spot for researchers and Australian regulators.

3. Ben Popkin 2018 ‘As algorithms take over, YouTube’s recommendations highlight a human problem’ *NBC News* <https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n867596>

4. Ralph Housego & Rys Farthing 2022 ‘Social Grooming’ *AQ Magazine* <https://www.jstor.org/stable/27161413>

5. Reset.Tech Australia & ISD 2022 *Algorithms as a weapon against women* <https://au.reset.tech/uploads/algorithms-as-a-weapon-against-women-reset-australia.pdf>

6. Chico Q. Camargo 2020 ‘YouTube’s algorithms might radicalise people – but the real problem is we’ve no idea how they work’ *The Conversation* <https://theconversation.com/youtubes-algorithms-might-radicalise-people-but-the-real-problem-is-weve-no-idea-how-they-work-129955>, and Brandy Zadrozny 2021 ‘Carol’s Journey’: What Facebook knew about how it radicalized users’ *NBC News* <https://www.nbcnews.com/tech/tech-news/facebook-knew-radicalized-users-rcna3581>

7. Richard Lawler 2022 ‘Meta reportedly plans to shut down CrowdTangle, its tool that tracks popular social media posts’ *The Verge* <https://www.theverge.com/2022/6/23/23180357/meta-crowdtangle-shut-down-facebook-misinformation-viral-news-tracker>



This research explores the effect of social media algorithms on political content promotion concerning the Voice referendum in Australia.

Recommender systems can also cause other harm, especially to children. Documented harms include promoting content related to eating disorders⁸ or fostering connections between children and adults' accounts, which creates contact risks.⁹

The harms of this can be significant. Last year, a UK coroner ruled that online content – promoted through content recommender systems – played more than a minor role in causing the suicide of a 14-year-old girl. The coroner concluded that she 'died from an act of self-harm while suffering from depression and the negative effects of online content'.¹⁰

8. Reset.Tech 2022 *Designing for Disorder*

<https://au.reset.tech/news/designing-for-disorder-instagram-s-pro-eating-disorder-bubble-in-australia/>

9. See for example, Australian Child Rights Taskforce 2023 *Letter to the eSafety Commissioner*

https://childrightstaskforce.org.au/wp-content/uploads/2023/01/Online-Safety-Codes_-_ACRT-letter-to-eSafety.pdf

10. BBC 2022 'Molly Russell inquest: Father makes social media plea' *BBC*

<https://www.bbc.com/news/uk-england-london-63073489>

Method

This research aimed to explore how the recommender systems of platforms can create divided political filter bubbles on two platforms: TikTok and X (formerly Twitter). These platforms were selected as examples, as existing research suggests that rabbit holes and filter bubbles also exist on other platforms, including Facebook¹¹ and YouTube.¹²

1. Establishing sock puppet accounts.

We established pairs of sock puppet accounts on TikTok and X. These sock puppet accounts were all notionally set to be 17 years old, Australian, set to private, and were clearly identified as research accounts.

2. Priming sock puppets.

Each sock puppet account was primed to be interested in content related to the Voice.

- **On TikTok**, the accounts were primed by watching 30 videos about the Voice referendum twice. These videos were neutral, i.e., neither leaning towards 'Yes' nor 'No'. This content mainly consisted of initial pieces about the Voice referendum, such as news publishers explaining the proposal for a Voice to Parliament, how to enrol, and announcements that parliament had voted to hold the referendum.
- **On X**, the accounts were primed by liking 15 neutral news publishers that were recommended to each account upon creation. These ranged from Channel 10 and Channel 7 to The Australian, Sky News, and The Guardian. After the initial 5 Xs, each account was recommended three additional accounts to follow, of which they followed two. Given the prominence

of the Voice in Australian news coverage at the time, this was sufficient to prime the accounts.

3. Training sock puppets.

Once these accounts were primed, we trained them and monitored the content that the recommender system served on the 'For You' Feed. Half of the accounts were conditioned to receive 'Yes' content recommendations, and the other half, 'No' content. Thus, whenever a sock puppet account encountered 'Yes' content, it watched it twice on TikTok or liked it on platform X. The corresponding account did the same for 'No' content.

- **On TikTok**, videos were watched in the order they were served on the For You Feed. Depending on the account, videos that were 'Yes' or 'No' were watched twice to train the algorithm.
- **On X**, Xs were read chronologically on the For You Feed. Depending on the account, Xs that were 'Yes' or 'No' were liked. The 'For You Feed' was refreshed frequently, every 10 Xs initially, then every 5 Xs towards the end of the experiment.

4. Analysing findings.

We analysed the rate and speed at which each account fell into a 'Yes' or a 'No' filter bubble.

¹¹. See Dick Lilly 2018, 'Social media's algorithms lead us down dark, divisive rabbit holes,' *The Seattle Times* <https://www.seattletimes.com/opinion/social-medias-algorithms-lead-us-down-dark-divisive-rabbit-holes/>; and Molly Killeen 2021, 'Facebook can lead users down vaccine misinformation rabbit hole' *EURACTIV*, Trust Project <https://www.euractiv.com/section/disinformation/news/facebook-can-lead-users-down-vaccine-misinformation-rabbit-hole/>.

¹². See Manoel Horta Ribeiro et al. 2020, 'Auditing radicalization pathways on YouTube' *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* <https://dl.acm.org/doi/abs/10.1145/3351095.3372879>; Tech Transparency Project 2021, 'YouTube's Filter Bubble Problem is Worse for Fox News Viewers' *Tech Transparency Project* <https://www.techtransparencyproject.org/articles/youtubes-filter-bubble-problem-worse-fox-news-viewers>; and Jonas Kaiser and Adrian Rauchfleisch 2020, 'How YouTube helps form homogeneous online communities' *Brookings Institution* <https://www.brookings.edu/articles/how-youtube-helps-form-homogeneous-online-communities/>. Note, the strength of the effect of YouTube's rabbit holes has been contested. See for example, Mark Ledwich and Anna Zaitsev 2020, 'Algorithmic extremism: Examining YouTube's rabbit hole of radicalization' *First Monday* <https://firstmonday.org/ojs/index.php/fm/article/view/10419>; Matus Tomlein et al. 2021, 'An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behaviour Changes' *Proceedings of the 15th ACM Conference on Recommender Systems* <https://dl.acm.org/doi/10.1145/3460231.3474241>; and Kaitlyn Tiffany 2023, 'Very, Very Few People Are Falling Down the YouTube Rabbit Hole' *The Atlantic* <https://www.theatlantic.com/technology/archive/2023/08/youtube-rabbit-holes-american-politics/675186/>.

Findings



We ran two pairs of sock puppet accounts on TikTok, thereby having two accounts trained to fall into a 'Yes' filter bubble and two accounts trained to fall into a 'No' filter bubble.

One of our accounts quickly fell into a 'Yes' filter bubble, beginning at about 50 video posts after priming. The filter bubble was exponential, indicating that the rate at which this account was being recommended 'Yes' content was increasing in acceleration. We halted this account at 200 videos as the filter bubble became apparent. The second account appeared to initially fall into a filter bubble; however, the recommender system

subsequently pulled it out. We halted this account at 400 videos when it became clear it was not going to be pulled into an exponential filter bubble.

Figure 1 illustrates the number of videos served to each sock puppet account and the cumulative count of 'Yes' videos viewed. The exponential bubble becomes visible for the first account, demonstrating that as more videos were served to the account overall, the rate at which it was served 'Yes' videos increased. For the second account that failed to fall into a filter bubble, the rate at which it was served 'Yes' videos did not increase over time.

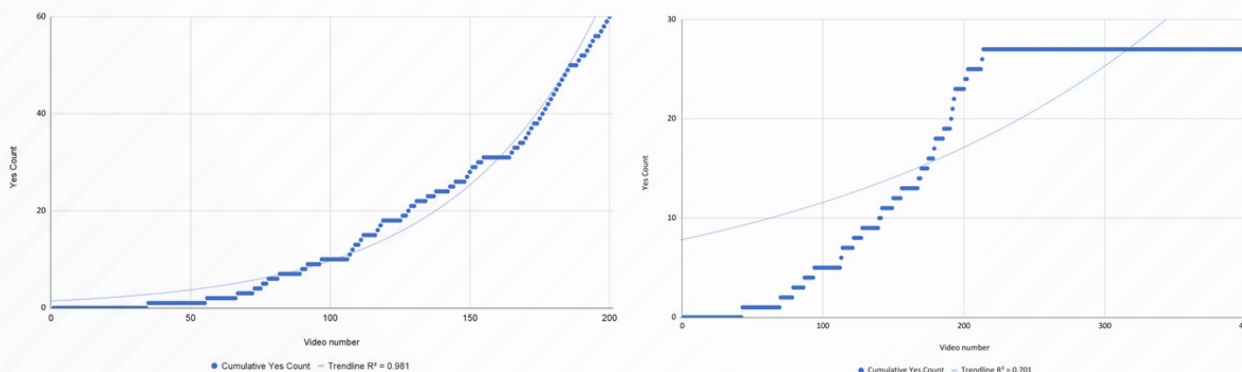


Figure 1: The two accounts that were trained to be recommended 'Yes' content.

Both of our 'No' sock puppets fell into exponential filter bubbles, although this process took longer. We 're-primed' the accounts twice by watching the same neutral content regarding the Voice when the recommender system had strayed into promoting celebrity content, for example.

At around 250 videos, both began falling into a 'No' filter bubble, and we ceased both accounts at 350–400 videos when it was clear that filter bubbles had emerged. Figure 2 visualises these filter bubbles, indicating that the rate at which accounts were served 'No' videos increases over time.

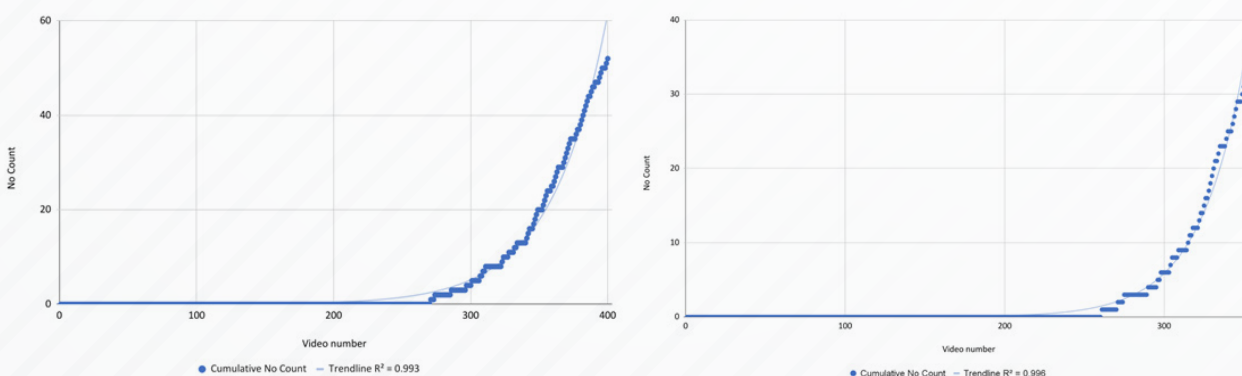


Figure 2: The two accounts that were trained to be recommended 'No' content.

X (Twitter)

We ran a pair of sock puppet accounts on X, with one account trained to fall into a 'Yes' rabbit hole and the other trained to fall into a 'No' rabbit hole.

The 'Yes'-trained account fell into a filter bubble aligned with 'Yes' content after approximately 200 Xs (tweets). 'Yes' related content started to appear straight after the sock puppet account was created and increased in frequency as the experiment went on. Figure 3 graphs the number of Xs served to the 'Yes' account and the cumulative count of 'Yes' Xs recommended, demonstrating that the rate at which the account is being recommended 'Yes'

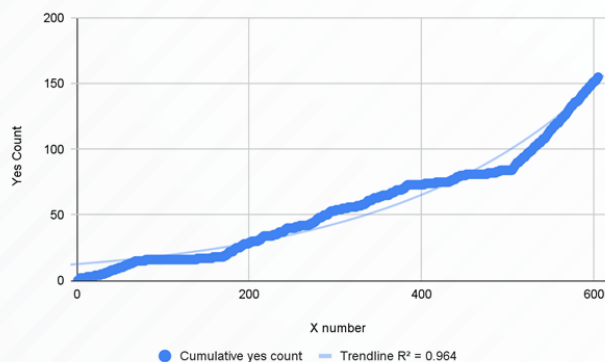


Figure 3: The account that was trained to be recommended 'Yes' content.

videos is increasing over time.

The 'No'-trained account fell into a filter bubble aligned with 'No' content. The recommender system pulled the account out of the filter bubble for around 100 Xs at approximately 200–300 pieces of content but then swiftly returned as the experiment progressed. Figure 4 graphs the number of Xs served to the 'No' sock puppet account and the cumulative count of 'No' Xs seen, demonstrating that the rate at which the account is being recommended 'No' content is increasing over time.

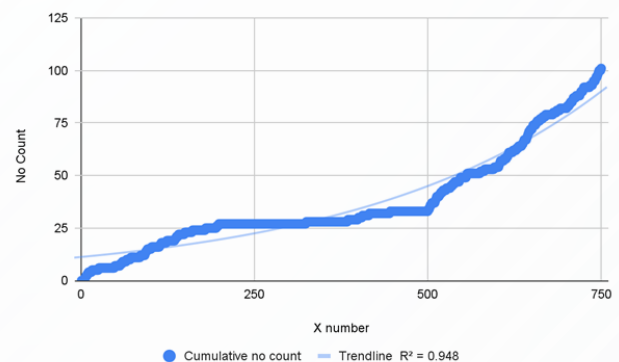


Figure 4: The account that was trained to be recommended 'No' content.

Conclusions

- This small-scale study highlights the importance of content recommender systems on platforms in promoting political content. Accounts that initially seek neutral Voice content could quite quickly fall into partisan filter bubbles, responding to the divisive actions of users.
- In many ways, this exemplifies recommender systems functioning as designed, which does not necessarily represent a problem in itself. This study does not make any claims about the effects of this on the political perspectives of social media users in Australia; rather, it raises questions about how the content recommender systems of these platforms may be shaping and dividing political discourse.
- Despite their potential impact on political debate, Australia's content recommender systems and algorithms operate 'in the dark'. Only the platforms truly know how their algorithms shape Australian political discourse.
- Other jurisdictions have implemented regulations mandating transparency through 'researcher access' to public interest data. As Australia considers its approach to digital regulation via the *Exposure Draft Communications Legislation Amendment (Combating Misinformation and Disinformation) Bill* and the review of the *Online Safety Act (2021)*, ensuring access to public-interest data should be a key priority.

