

Inquiry into the Impact of Social Media on Elections & Electoral Administration

Written submission to the Electoral Matters Committee, Parliament of Victoria, Australia
27 August 2019

Authors:

[Vian Bakir](#), Professor of Journalism & Political Communication, Bangor University, Wales, UK.

[Andrew McStay](#), Professor of Digital Life, Bangor University, Wales, UK.

Both from [Network for Study of Media & Persuasive Communication](#), and the [Emotional AI Lab](#)

Terms of reference

This submission addresses the following areas:¹

- How is social media changing elections?
- Social media and online political advertising: problems and solutions.
- Quick, uncontroversial fixes.
- Longer-term investigation required.

This submission ignores many of the complexities regarding the interplay of social media and elections. It does not, for instance, address issues of digital literacy, foreign influence campaigns, and the declining health of the wider news media landscape. Rather, *this submission focuses on online political advertising*, in order to address the core questions posed by this inquiry.

Summary

1. Social media campaigning could change elections for the better by increasing engagement with citizens on issues that they care about. However, it proliferates microtargeted disinformation, which generates many democratic harms.
2. Online advertising is a key economic driver behind the proliferation of online disinformation. This has manifested in two main ways.
 - a) The ecology of behavioural advertising funds fake news sites through use of adtech to profile and target people.
 - b) Datafied emotional content is optimised to generate Facebook shares for internet traffic and advertising income (clickbait audiences): this takes form across fake and real news outlets (e.g. via misleading headlines), in hate speech (e.g. via inflammatory and discriminatory ads) and in political campaigning (e.g. via political ads and issues based ads).
3. Dominant digital platforms have taken various actions in recent years to disrupt the business models for the production and amplification of microtargeted disinformation. Governments have also taken certain actions, and civil society actors have recommended many more.
4. There are two quick, uncontroversial fixes that could be taken in Australia.
 - a) *Increase the transparency of online political ads* by ensuring that social media platforms that allow political ads and political issue ads keep *accessible, searchable libraries* of these ads.
 - b) Ensure that there are *digital imprints* to hold advertisers accountable for what they say.
5. Longer term, close attention should be paid to the legality of microtargeting. Also, more studies are needed into the impact of microtargeting on electorates, both in terms of direct influence and regarding wider trust in the fair conduct of elections.

¹ Funding: This work is supported by the UK Economic & Social Research Council, Grant number: ES/T00696X/1.

How is social media changing elections?

For Better ...

On the positive side, social media campaigning may *increase engagement with citizens on issues that they care about*.² Where conducted openly and honestly, microtargeting can help manifest voters' desires, concerns and policy preferences to politicians, thereby helping elected leaders develop programs that meet voters' needs.³ For instance, the UK's Electoral Commission (the elections regulator) observes that there is value in microtargeting if conducted transparently and honestly: 'For example, people who live three streets from the hospital that they are particularly interested in will want to know about plans to close that hospital or suchlike.'⁴

For Worse ...

On the negative side, social media campaigning generates multiple harms. It (1) produces wrongly informed citizens, that (2) may stay wrongly informed in digital echo chambers (especially during periods of intensive efforts to persuade and influence, such as elections and referenda) and (3) be emotionally provoked given the affective nature of much online disinformation. Added to this is a fourth problem: (4) contagion, where false information spreads across social media and mainstream news. Furthermore, (5) microtargeting enables a range of democratic harms, from fragmentation of important national conversations through to undue influence over susceptible citizens. Related (6) is the impact of the disinformation media ecology on seeding distrust in democratic institutions and processes. These harms are described below.

Harm 1. Wrongly Informed Citizens

Big data studies show that fake news on social media is a significant pollutant during elections in swing states⁵ and more generally.⁶ Presenting a more mixed picture, a study of trends in diffusion of content across Facebook and Twitter across 2015–18 finds that user interactions with false content rose steadily on both Facebook and Twitter through the end of 2016, but since then, interactions with false content have fallen sharply on Facebook while continuing to rise on Twitter.⁷ A study of the 2019 UK General Election found that a significant quantity of adverts from the political parties contained misleading statements.⁸ *Widespread recirculation of deceptive and false content posing as news does not bode well for the factual foundations on which citizens form opinions, and the nation's consequent democratic health.*

Harm 2: Digital Echo Chambers: Remaining Wrongly Informed

Echo chambers exist where information, ideas or beliefs are amplified and reinforced by communication and repetition inside a defined system where competing views are under-represented.⁹ Algorithmically-created echo chambers, or *filter bubbles*, arise when algorithms

² Information Commissioners Office. 2018. *Investigation into the use of data analytics in political campaigns: A report to Parliament*, 6 November 2018. <https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf>

³ Perloff, R.M. 2018. *The Dynamics of Political Communication: Media and Politics in a Digital Age*. New York and London: Routledge. P.250.

⁴ Bassett, C. 2018. Oral evidence: Fake News, HC 363, 2018 15 May: 34-35.

<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/82877.html>

⁵ Howard, P.N., Kollanyi, B., Bradshaw, S., & Neudert, L.M. 2017. Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States? *COMPROP DATA MEMO*, 28 September. <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/09/Polarizing-Content-and-Swing-States.pdf>

⁶ Vargo, C.J., Guo, L. & Amazeen, M.A. 2018. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 20(5), 20282049. DOI: 10.1177/14614481712086.

⁷ Allcott, H., Gentzkow, M. & Yu, C. 2019. Trends in the diffusion of misinformation on social media. *Research and Politics*, April-June: 18. DOI: 10.1177/205316801948554

⁸ Full Fact. 2019. The facts behind Labour and Conservative Facebook ads in this election, 11 Dec. <https://fullfact.org/election-2019/ads/>

⁹ Sunstein, C.R. 2002. The Law of Group Polarization. *The Journal of Political Philosophy*, 10(2): 175-195.

<https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-9760.00148>.

applied to online content selectively gauge what information a user wants to see based on information about the user, their connections, browsing history, purchases, and what they post and search. This results in users becoming separated from exposure to wider information that disagrees with their views.¹⁰

Big data studies confirm that filter bubbles exist, but there remains disagreement on the implications of filter bubbles for norms of behaviour important to democracy. For instance, a study of the US press and social media landscape in the 18 months leading up to the US presidential election (2015-16) shows that the right-wing media ecosystem was more insulated than the left-wing media ecosystem, and so was susceptible to disinformation.¹¹

In terms of negative implications for democracy, many studies find that these filter bubbles result in limited exposure to, and lack of engagement with, different ideas and other people's viewpoints on social media (especially Facebook).¹² This fuels conspiracy theories¹³ and the spread of unverified rumours and fake news;¹⁴ and makes users resistant to debunking of disinformation circulating within the filter bubbles.¹⁵ This state of affairs is problematic for democracy, because, to make informed decisions, all citizens need access to, and engagement with, a sufficiently diverse body of information about public life.

Conversely, on the positive side, studies also show that social networks lead to greater exposure to diverse ideas.¹⁶ Studies on news exposure on multiple platforms, find that the effect of personalisation is smaller than often assumed; and that recommendation algorithms *increase* rather than decrease diversity in news consumption.¹⁷

On balance, we suggest that filter bubbles are likely to be most problematic during periods of intensive efforts to persuade and influence – such as general elections and referenda. Furthermore, the rise in targeting and profiling of audiences during these periods means that individual users can be exposed to highly customised campaign messages. This could lead to people having very different experiences, and presumably understanding, of what should be a common national campaign.

¹⁰ Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press.

¹¹ Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E. & Benkler, Y. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. August 16. Berkman Klein Center for Internet and Society at Harvard University. <https://cyber.harvard.edu/publications/2017/08/mediacloud>

¹² Bessi, A., Petroni F., Del Vicario, M., Zollo F., Anagnostopoulos A., Scala A., Caldarelli G. & Quattrociocchi, W. 2016. Homophily and polarization in the age of misinformation. *Eur. Phys. J. Special Topics*, 225: 2047-2059. DOI: 10.1140/epjst/e2015-50319-0. Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G. & Quattrociocchi, W. 2015. Science vs conspiracy: collective narratives in the age of misinformation. *PLoS one*, 10(2): e0118093. <https://doi.org/10.1371/journal.pone.0118093>. Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarella, G., Stanley, H. E. & Quattrociocchi, W. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3): 554-559. <http://www.pnas.org/content/113/3/554.full.pdf>. Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E. and Benkler, Y. 2017. *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. August 16. Berkman Klein Center for Internet and Society at Harvard University. <https://cyber.harvard.edu/publications/2017/08/mediacloud>. Milani, E., Weitkamp, E. & Webb, P. 2020. The Visual Vaccine Debate on Twitter: A Social Network Analysis. *Media and Communication*, 8, 2: 364-375. DOI: 10.17645/mac.v8i2.2847 www.cogitatiopress.com/mediaandcommunication/article/view/2847. Quattrociocchi, W., Scala, A. & Sunstein, C.R. 2016. "Echo Chambers on Facebook." SSRN. <https://papers.ssrn.com/sol3/papers.cfm?abstractid=2795110>. Zollo, F. et al. 2017. Debunking in a world of tribes. *PLoS ONE* 12(7): e0181821. <https://doi.org/10.1371/journal.pone.0181821>.

¹³ Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G. & Quattrociocchi, W. 2015. Science vs conspiracy: collective narratives in the age of misinformation. *PLoS one*, 10(2): e0118093. <https://doi.org/10.1371/journal.pone.0118093>

¹⁴ Bessi, A., Scala, A., Rossi, L., Zhang, Q. & Quattrociocchi, W. 2014. The economy of attention in the age of (mis) information. *Journal of Trust Management*, 1(12). DOI 10.1186/s40493-014-0012-y. Bessi, A., Petroni F., Del Vicario, M., Zollo F., Anagnostopoulos A., Scala A., Caldarelli G. & Quattrociocchi, W. 2016. Homophily and polarization in the age of misinformation. *Eur. Phys. J. Special Topics*, 225: 2047-2059. DOI: 10.1140/epjst/e2015-50319-0. Brummette, J., DiStaso, M., Vafeiadis, M. & Messner, M. 2018. Read All About It: The Politicization of "Fake News" on Twitter. *Journalism & Mass Communication Quarterly*, 95(2) 497-517. <https://doi.org/10.1177/1077699018769906>.

¹⁵ Zollo, F. et al. 2017. Debunking in a world of tribes. *PLoS ONE* 12(7): e0181821. <https://doi.org/10.1371/journal.pone.0181821>

¹⁶ Flaxman, S., Goel, S. & Rao, J.M. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(1): 298-320. doi:10.1093/poq/nfw006. Messing, S. & Westwood, S.J. 2012. Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research* 41:1042-1063. DOI: 10.1177/0093650212466406.

¹⁷ Nechushtai, E. & Lewis, S.C. 2019. What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, 90: 298-307. <https://doi.org/10.1016/j.chb.2018.07.043>.

Harm 3: Affective Content, Incivility, Polarisation

The deliberately *affective* nature of the contemporary disinformation media ecology drives *tribal, partisan polarisation*,¹⁸ and may also *fuel the spread of disinformation* where people share fake news as an expression of belonging to a specific social group rather than an act of rational information sharing.¹⁹ It also drives *uncivil behaviour*, although evidence of the impacts on political mobilisation are mixed. While uncivil behaviour can increase political mobilisation and voter turnout²⁰ it can also have the opposite effect, eroding political trust, decreasing the perceived legitimacy of political figures, and reducing political participation intention and policy support.²¹

Recognising its central role in driving polarisation, Facebook has taken some corrective measures. These include:

- Recalibrating its News Feed ranking in 2018 to prioritise posts from friends and family over news content. Facebook also reports reducing its clickbait headlines, reducing its links to spam and misleading posts, and improving its comment rankings to show people higher quality information.
- Building its Global Integrity Team across the last four years to address safety and security issues, including polarisation.
- Restricting Recommendations from its Pages and Groups if they repeatedly share content that violates Facebook's Community Standards or is rated false by fact-checkers.

However, as of 2020, civil rights auditors of Facebook's US app do not believe that Facebook is sufficiently attuned to the depth of concern on the issue of polarisation and the way that the algorithms used by Facebook fuel extreme and polarising content (even with the above measures).

22

Harm 4: Contagion

In 2012, Facebook demonstrated that *emotions are contagious on its platform*.²³ Since then, studies have demonstrated that *false information (especially false political news) is contagious online*,²⁴ influencing mainstream news and wider social media,²⁵ thereby spreading its pollutants far and wide.

Harm 5: Microtargeting Harms, from Fragmentation to Undue Influence

Digital microtargeting provides campaigns with *new, opaque, poorly understood ways to study and mobilise potential voters, raising classic questions of covert manipulation and media effects*. This gives rise to the following five microtargeting harms.

- a) Fragmentation of national conversations.** Microtargeting weakens our ability to hold political campaigners, and those subsequently elected, to account. If false information, or deception takes place via microtargeting, and if this is not publicly scrutinised, then there is little chance of those elected on such platforms being held to public account. This risk

¹⁸ Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., & Quattrocchi, W. 2016. Echo Chambers: Emotional Contagion and Group Polarization on Facebook. *Physics and Society*, <https://arxiv.org/pdf/1607.01032.pdf>. Hong, S. & Kim, S.H. 2016. 'Political polarisation on Twitter: Implications for the use of social media in digital governments', *Government Information Quarterly*. Spohr, D. 2017. Fake news and ideological polarisation: Filter bubbles and selective exposure on social media, *Business Information Review*, 34(3), pp. 150-160.

¹⁹ Polletta, F. & Callahan, J. 2017. Deep stories, nostalgia narratives, and fake news: Storytelling in the Trump era. *American Journal of Cultural Sociology*, 5(3), 392-408. <https://doi.org/10.1057/s41290-017-0037-7>

²⁰ Lu, Y. & Myrick, J.G. 2016. Cross-cutting exposure on Facebook and political participation: Unravelling the effects of emotional responses and online incivility. *Journal of Media Psychology: Theories, Methods, and Applications*, 28(3), 100-110. <http://dx.doi.org/10.1027/1864-1105/a000203>

²¹ Fridkin, K.L. & Kenney, P. 2008. The dimensions of negative messages. *American Politics Research*, 36, 694-723. Otto, L.P., Lecheler, S. & Schuck, A.R.T. 2019. Is Context the Key? The (Non-) Differential Effects of Mediated Incivility in Three European Countries, *Political Communication*, DOI: 10.1080/10584609.2019.1663324

²² Murphy, L.W. 2020. *Facebook's Civil Rights Audit - Final Report*, 8 July. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

²³ Kramer, A.D.I., Guillory, J.E. & Hancock, J.T. 2014. Experimental evidence of massive-scale emotional contagion through social networks, *Proceedings of the National Academy of Sciences*, 111 (29): 8788-90.

²⁴ Vosoughi, S., Roy, D. & Aral, S. 2018. The spread of true and false news online, *Science*, 359 (6380): 1146-1151.

²⁵ Benkler, Y., Faris, R., Roberts, H. & Zuckerman, E. 2017. Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review*. <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>

intensifies as algorithmic marketing techniques become available to all political parties, enabling parties to routinely run millions of algorithmically tuned messages, on a scale that could overwhelm regulators. Already, Facebook's tool, Dynamic Creative, allows campaigns to use predefined design features of an ad and machine learning to construct thousands of variations of the ad, present them to users, and find optimal combinations based on engagement metrics.²⁶

b) Targeted suppression of voters and civic activity via Facebook. Political behaviour change and data analytics companies (such as the now defunct Cambridge Analytica/SCL Group) offer voter suppression as part of their services.²⁷ In the 2016 Trump presidential campaign, Brad Parscale, the campaign's digital director in 2016 (and also for 2020), used Facebook's Lookalike Audiences ad tool to expand the number of people the campaign could target by identifying voters who were not Trump supporters, and then targeting them with psychographic messaging designed to discourage them from voting.²⁸

c) Opaque exploitation of vulnerabilities via Facebook and data analytics companies. This encompasses targeting people who are not equipped to evaluate the message. Cambridge Analytica whistleblower, Christopher Wylie, explained in 2018 that the company was built on exploiting Facebook to harvest millions of people's profiles to then build models to 'target their inner demons'²⁹ including targeting people prone to adopting conspiracies with conspiratorial messaging, whether through adverts, blogs or fake news.³⁰ Tristan Harris (Co-founder and Executive Director, Center for Humane Technology) demands a ban on microtargeted advertising because:

there is no way in which Facebook can control the exponentially complex set of things that people let advertisers enter into the platforms to target users' baseline. For example, basically engineers only speak so many languages and they are stirring the pot of 2 billion people who have languages that they don't even speak, including in very emergent sensitive markets where they have just got the internet like in Myanmar and Sri Lanka. It is possible for people to create culture wars in those languages because there is much less oversight by Facebook and different parties.³¹

d) Opaque and inaccurate profiling via political parties. While there has been much attention to the activities of social media companies in political profiling, there has been much less on the role of political parties who commission such activities and process large quantities of personal data. An exception is a study by Open Rights Group (ORG) into how political parties use the personal data of UK voters. It finds *extensive, but unjustifiable, use of personal data to try and profile people, which is then used to try to tailor a political relationship with that person*. It finds that political parties had a confused understanding of data protection law: they used the justification of 'democratic engagement' for processing, whereas this should 'only really be used to justify use of the electoral register - not to use or generate information about a person's income, nationality, or political opinions'.³² On the

²⁶ Bartlett, J., Smith, J., & Acton, R. 2018. *The Future of Political Campaigning*. Demos, p.39. <https://ico.org.uk/media/2259365/the-future-of-political-campaigning.pdf>

²⁷ Bakir, V. 2020. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Political Communication*. 5(67). doi: 10.3389/fcomm.2020.00067

²⁸ Green, J., & Issenberg, S. 2016. Inside the Trump Bunker, with days to go, *Bloomberg*. <https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go>

²⁹ Cadwalladr, C. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 17 March. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

³⁰ Wylie, C. 2018. Oral evidence: Fake News, HC 363, 27 March, pp. 21-22. <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/81022.pdf>

³¹ Harris, T. 2018. Oral evidence: Fake News, HC 363. 2018 22 May, p.10. <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/disinformation-and-fake-news/oral/83304.html>

³² Crowe, P., Rice, M. & Santi, M.D. 2020. *Who Do They Think We Are? Political Parties, Political Profiling, and the Law*. ORG, 23 June. P.4. <https://www.openrightsgroup.org/publications/who-do-they-think-we-are-report/>

basis of their study of Data Subject Access Request responses from UK political parties, ORG conclude that *profiling by UK political parties is mostly inaccurate*: ‘despite all the hype, profiling by political parties is more Mr Bean than Machiavelli’.³³

e) **Undue Influence?** Arguably, to avoid undue influence, communications that try to persuade should be guided by the principle of informed consent, so that the persuadee’s decision to accept a particular viewpoint is both informed and freely chosen. However, this is disabled by deception and coercion.³⁴ Deception can take many forms, such as fake news,³⁵ lying,³⁶ omission,³⁷ misdirection³⁸ and misleading clickbait.³⁹ Coercion can take many forms such as application of force (threats, violence); deliberate limitation of people’s choices by making them feel or behave a certain way; and preclusion of reflection and deliberation. Manipulation of the digital media ecology is coercive (choice limiting) where optimised microtargeting is used to suppress a target audience’s desire to vote, or to modulate its exposure to information that is important in informing voting choices and behaviour.⁴⁰

Such manipulation would go against the principle of free and fair elections.⁴¹ Moreover, approaching this problem from the perspective of legal rights, if profiling and microtargeting is used to surreptitiously influence people’s thought, then this may contravene the right to freedom of thought - an absolute right (i.e. no restrictions are allowed) in international human rights law since 1976, enshrined in Article 18 of the International Covenant on Civil and Political Rights.⁴² Freedom of thought includes the freedom to keep our thoughts private so that we may not be coerced into revealing them; and freedom from indoctrination or influence on our conscious or subconscious mind through manipulation. However, this fundamental right has received little attention in the courts or in academic analysis, partly because of an assumption that our inner thoughts were beyond reach, and partly because of an absence of precise definition as to what constitutes thought.⁴³

Existing research into the effects of social media manipulation has largely focused on measuring its purported impact on opinion swings and voting behaviour, and has so far concluded that impacts are minimal.⁴⁴ However, proving influence, is hard: we need more user studies, to examine to what extent different groups are subjected to profiled, choice-

³³ Crowe, P., Rice, M. & Santi, M.D. 2020. *Who Do They Think We Are? Political Parties, Political Profiling, and the Law*. ORG, 23 June. P.9.

<https://www.openrightsgroup.org/publications/who-do-they-think-we-are-report/>

³⁴ Bakir, V., Herring, E., Miller, D., & Robinson, P. 2019. Organized persuasive communication: a conceptual framework. *Critical Sociology* 45(3).

<https://doi.org/10.1177/0896920518764586>

³⁵ Egelhofer, J.L. & Lecheler, S. 2019. Fake news as a two-dimensional phenomenon: a framework and research agenda, *Annals of the International Communication Association*, 43(2), 97-116, DOI: 10.1080/23808985.2019.1602782.

³⁶ Bakir, V., Herring, E., Miller, D. & Robinson, P. 2018. Lying and Deception in Politics. In Meibauer, J. (ed.) *The Oxford Handbook of Politics and Lying*. Oxford Univ. Press.

³⁷ Bakir, V., Herring, E., Miller, D., & Robinson, P. 2019. Organized persuasive communication: a conceptual framework. *Critical Sociology* 45(3).

<https://doi.org/10.1177/0896920518764586>

³⁸ Bakir, V. 2017. Political-Intelligence Elites, Strategic Political Communication and the Press: the Need for, and Utility of, a Benchmark of Public Accountability Demands. *Intelligence and National Security*, 32(3). <https://doi.org/10.1080/02684527.2016.1231866>

³⁹ Bakir, V. & McStay, A. 2020. Empathic Media, Emotional AI and the Optimization of Disinformation. In M. Bolter and E. Davis (eds.), *Affective Politics of Digital Media: Propaganda by Other Means*. New York: Routledge.

⁴⁰ Bakir, V. 2020. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica’s Psychographic Profiling and Targeting. *Frontiers in Political Communication*. 5(67). doi: 10.3389/fcomm.2020.00067

⁴¹ Marchal, N. 2020. Conceptualizing the impact of digital interference in elections: a framework and agenda for future research. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.3536281>

⁴² International Covenant on Civil and Political Rights. 1976. <https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx>

⁴³ Alegre, S. 2017. Opinion. Rethinking Freedom of Thought for the 21st Century. *European Human Rights Law Review*, 3: 221-233.

<https://www.doughtystreet.co.uk/sites/default/files/media/document/Rethinking%20Freedom%20of%20Thought%20for%20the%2021st.pdf>. McCarthy-Jones, S. 2019. The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century. *Frontiers in Artificial Intelligence*, 2(19): 1-17.

<https://doi.org/10.3389/frai.2019.00019>

⁴⁴ Marchal, N. 2020. Conceptualizing the impact of digital interference in elections: a framework and agenda for future research. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.3536281>

limiting coercion during voting periods, such as whether they are exposed to a significant stream of targeted information designed to change their behaviour.⁴⁵ *The nature and extent of undue influence, then, requires greater conceptualisation and evidencing, both in user-based and legal studies.*

Harm 6: Seeding Distrust in Democratic Institutions and Processes.

It is important for electorates to have fair access to information about parties and candidates that are free from undue influence or coercion in order to prevent damaging effects on wider democratic processes such as trust in institutions. Regardless of whether people's political viewpoints are directly influenced by profiling and targeting, *opaque attempts to manipulate them may undermine popular perceptions of electoral fairness and integrity with potentially far-reaching consequences for public trust.*⁴⁶ Indeed, the study by ORG into how political parties use the personal data of UK voters concludes that opaque data profiling practices leaves most voters unaware of what political parties do with their personal data, which has the potential to seriously undermine trust in the democratic process and damage its integrity.⁴⁷ Certainly, surveys in multiple countries (including Australia) show that *many people are unaware of, and/or dislike, efforts by political campaigners to provide more efficient ways to influence human behaviour by using opaque data science methods.*⁴⁸

⁴⁵ Bakir, V. 2020. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Political Communication*. 5(67). doi: 10.3389/fcomm.2020.00067

⁴⁶ Marchal, N. 2020. Conceptualizing the impact of digital interference in elections: a framework and agenda for future research. *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.3536281>

⁴⁷ Crowe, P., Rice, M. & Santi, M.D. 2020. *Who Do They Think We Are? Political Parties, Political Profiling, and the Law*. ORG, 23 June. P.4. <https://www.openrightsgroup.org/publications/who-do-they-think-we-are-report/>

⁴⁸ Bakir, V. 2020. Psychological Operations in Digital Political Campaigns: Assessing Cambridge Analytica's Psychographic Profiling and Targeting. *Frontiers in Political Communication*. 5(67). doi: 10.3389/fcomm.2020.00067. Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. 2020. *Reuters Institute digital news report 2020*. P.42. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Social Media & Online Political Advertising: Problems & Solutions

As noted above, online disinformation inflicts multiple potential harms upon democracies. Online advertising is a key economic driver behind the proliferation of online disinformation generally. This manifests in two key ways.

- The ecology of behavioural advertising funds fake news sites through the use of adtech to profile and target people.
- Datafied emotional content is optimised to generate *Facebook* shares for internet traffic and advertising income (clickbait audiences): this takes form across fake and real news outlets (e.g. via misleading headlines), in hate speech (e.g. via inflammatory and discriminatory ads) and in political campaigning (e.g. via political ads and issues based ads).

Dominant digital platforms have taken various actions in recent years to disrupt the business models for their production and amplification of disinformation, and these are examined below.

Funding Fake News sites on Google via Adtech

It is instructive to look at the USA, as social media has been used there to spread emotive disinformation to manipulate elections since 2006. In 2006 political blogs tried to influence US elections by gaming search engines, pushing web pages carrying negative content to the top of relevant search results. Using link bombing techniques ('Googlebombing'), website masters and bloggers used the anchor text to associate an obscure, negative term with a public entity. During the 2006 US midterm congressional election, an effort to manipulate ranking results to promote negative stories about Republican incumbents running for Congress took place openly solicited by the liberal blog, MyDD.com (My Direct Democracy). Google took steps to curb such activity by promoting uncontroversial results in the first page, and it was found that political spammers were not very successful in the 2008 Congressional elections. However, an examination of political community behaviour on Twitter in the Massachusetts Senate race between Martha Coakley and Scott Brown in 2010 found a 'Twitter-bomb', namely, an organised effort to spread misinformation about democratic candidate Coakley through anonymous Twitter accounts: the success of a Twitter-bomb relies on targeting users interested in the spam topic, and relying on those users to spread the spam.⁴⁹

By 2016, social media played a vital role in Donald Trump's presidential campaign. Tweets from Trump and his supporters were more likely to go viral than those in support of his rival, Hillary Clinton.⁵⁰ Trump won Facebook ad auctions due to the likely engagement his content would generate compared with Clinton, taking shape in racism, misogyny and an antagonistic identity.⁵¹ Similarly, infiltration was successfully used on Facebook to spread fake news for financial benefit through online advertising.⁵² Some of these 100% fake news stories originated in the USA but many were fuelled by computer science students in Veles, Macedonia who launched multiple American politics websites carrying fake news stories. Analysis by journalists suggests that such websites were set up to make money rather than spread propaganda.⁵³

Google has since taken actions to address those parts of the disinformation media ecology fuelled by its advertising model. (In 2019 Google made US\$134.91 billion of its overall US\$160 billion annual revenue through adverts.⁵⁴) In Google's written submission to the UK Parliamentary

⁴⁹ Mustafaraj, E. & Metaxas, P.T. 2017. The fake news spreading plague: Was it preventable? <http://arxiv.org/abs/1703.06988>. Metaxas, P. T. & Mustafaraj, E. 2010. From obscurity to prominence in minutes: Political speech and real-time search. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, <https://repository.wellesley.edu/cgi/viewcontent.cgi?article=1008&context=computersciencefaculty>

⁵⁰ Rossman, S. 2017. Hillary Clinton lost the presidential Twitter war, study shows. *USA Today*, July 13.

<https://www.usatoday.com/story/news/politics/onpolitics/2017/07/13/hillary-clinton-lost-presidential-twitter-war-study-shows/473576001/>

⁵¹ Jutel, O. 2020. Digital Teleologies: Blockchain, Affect and Digital Teleologies. In Boler, Megan & Davis, Elizabeth (eds.) *Affective Politics of Digital Media: Propaganda by Other Means*. New York: Routledge, 101-115.

⁵² Mustafaraj, E. & Metaxas, P.T. 2017. The fake news spreading plague: Was it preventable? <http://arxiv.org/abs/1703.06988>

⁵³ Bakir, V. & McStay, A. 2020. Empathic Media, Emotional AI and the Optimization of Disinformation. In M. Boler and E. Davis (eds.), *Affective Politics of Digital Media: Propaganda by Other Means*. New York: Routledge.

⁵⁴ Clement, J. 2020. Google: annual advertising revenue 2001-2019. Statista. <https://www.statista.com/statistics/266249/advertising-revenue-of-google/>

Inquiry into Fake News and Disinformation, it outlines measures it had taken since November 2016 to help stifle fake news. These include new ad network policies against misrepresentation, which target website owners who misrepresent who they are and who deceive users with their content. Google reports that a month after it changed its ad network policy, it had identified 550 leads for sites suspected of misrepresenting content, including impersonating news organisations, and then removed nearly 200 publishers from their ad networks permanently.⁵⁵ Several months later, February 2017 surfaced one of the first major brand safety scandals, as ads from the world's biggest brands appeared on YouTube videos and websites of extremist and terrorist organisations via programmatic methods. The following month, as leading companies stopped spending ad money on YouTube, Google announced changes to YouTube that emphasised policies which refined definitions of hate speech or inflammatory content; enforced existing rules regarding takedowns of content; and enabled controls for advertisers to implement brand safety.⁵⁶

By March 2018, the European Commission was able to report that online platforms were tackling disinformation by disrupting the business model for its production and amplification ('follow-the-money'), via three aims:⁵⁷

- Advertising networks not placing ads on websites identified as purveyors of disinformation, thereby directly reducing income to disinformation providers.
- Advertising networks not disbursing revenues to sites and partners until they have been able to confirm that they operate within relevant terms and conditions.
- Advertising providers not accepting ads from disinformation sources and clearly marking political ads as sponsored content to create transparency.

Funding Hate Speech and Deceptive Political Ads on Facebook

Facebook makes nearly its entire US\$70 billion in annual revenue (2019) through digital advertising.⁵⁸ By 2019, the Final Report from the UK's Inquiry into Fake News and Disinformation notes that Facebook gives the impression of wanting to tackle disinformation (by, for instance, working with fact-checkers), but observes that: 'there is considerable obfuscation concerning the auditing of its adverts, which provide Facebook with its ever-increasing revenue. To make informed judgments about the adverts presented to them on Facebook, users need to see the source and purpose behind the content'.⁵⁹ Facebook has since taken measures to address some of the ills arising from its business model and popularity. These include many changes, but perhaps the most significant, given Facebook's business model, are changes to aspects of its advertising. This section examines what Facebook has done with its advertising policies to combat hate speech and political disinformation. It also highlights the inadequacies of these policies, and presents what other stakeholders have advocated or done to further address these issues.

Combating online hate speech. Facebook adopted an *Inflammatory Ads policy* in 2018, prohibiting certain types of attacks or fear-mongering claims made against people based on their race, religion, or other protected characteristics that would not otherwise be prohibited under

⁵⁵ Google. 2017. Written evidence submitted by Google (FNW0123). UK Parliamentary Inquiry into Fake News and Disinformation. Digital, Culture, Media and Sport Committee. <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/68824.html>

⁵⁶ Scott, S. 2020. Why this Facebook advertising boycott is different. *The Drum*, 8 July. <https://www.thedrum.com/opinion/2020/07/08/why-facebook-advertising-boycott-different>. Connelly, T. 2017. Google announces strategy to address concerns of ad misplacement and brand safety. *The Drum*, 20 March. <https://www.thedrum.com/news/2017/03/20/google-announces-strategy-address-concerns-ad-misplacement-and-brand-safety>

⁵⁷ European Commission. 2018. *A Multi-Dimensional Approach to Disinformation* (March), Directorate-General for Communication Networks, Content & Technology, p.15. <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>

⁵⁸ Clement, J. 2020. Facebook's annual revenue from 2009 to 2019. Statista. <https://www.statista.com/statistics/268604/annual-revenue-of-facebook/#:~:text=Facebook%3A%20annual%20revenue%202009%2D2019&text=In%202019%2C%20Facebook's%20revenue%20amounted,in%20the%20previous%20fiscal%20year>.

⁵⁹ DCMS. 2019. *Disinformation and 'fake news': Final Report*. 14 February. Digital, Culture, Media and Sport Committee, House of Commons 1791, p.85. Available at: <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/1791/1791.pdf>

Facebook's hate speech policies. The policy prohibits claims such as allegations that a racial group will 'take over' or that a religious or immigrant group as a whole represents a criminal threat. However, policies such as these have been criticised for being reactive and piecemeal, taking place only after negative civil rights impacts have already been felt.⁶⁰ For instance, Facebook has recently been accused of inadequately protecting and supporting black users, as well as excoriated for its long-standing failure to call out Holocaust denial as hate speech (rather than as misinformation). Across summer 2020, the killing of George Floyd during his arrest by US police, captured and widely shared on 'sousveillant' video,⁶¹ sparked Black Lives Matter protests across the world. Infamously, in May 2020, Facebook had allowed messages from US President Donald Trump that suggested that Black Lives Matter protestors could be shot if they continued to protest. One such message ends: 'These THUGS are dishonoring the memory of George Floyd, and I won't let that happen. Just spoke to Governor Tim Walz and told him that the Military is with him all the way. Any difficulty and we will assume control but, when the looting starts, the shooting starts'. A resulting 'Stop Hate for Profit' campaign with a month-long boycott of Facebook was co-organised by civil rights groups, producing the biggest boycott in Facebook's history by advertisers. (In contrast to Facebook, Twitter ruled that Trump's messages glorified violence: it kept the tweet up but placed a public interest notice on it and prevented users from replying or retweeting the post - the first time that Twitter had applied such a warning to President Trump's tweets.⁶²)

Combating deceptive online political ads. Finding solutions to online political ads that fuel the disinformation media ecology has hitherto proven difficult, for reasons ranging from understanding the opaque and microtargeted nature of digital political advertising through to defining what constitutes a 'political' ad. In June 2018, Facebook told the UK Parliamentary Inquiry into Fake News and Disinformation that that they had no reliable way of categorising which adverts could be classified as political; and that issue-based advertising is particularly difficult to define at a global scale as there is no universal definition of a political issue advert. To help address this, Facebook collaborated with external stakeholders and invested in machine learning to better assess adverts that fall into specific categories (like political and issues adverts) to help Facebook identify and enforce policies and tools that may apply.⁶³ By September 2019, Facebook had formulated policies to discourage voter suppression - its 'Don't Vote Ads Policy' and a 'Don't Participate in Census Ads Policy'. Together, these policies prohibit ads targeting the USA that portray voting or census and include threats that voting will result in adverse law enforcement consequences (e.g. arrest, deportation, imprisonment); and statements that encourage coordinated interference in elections. Facebook also adopted a new policy prohibiting threats of violence relating to voting, voter registration or the outcome of elections (e.g. statements to bring weapons to polling places).⁶⁴

It is not just Facebook that has struggled with how to define political ads. Google has still not defined what it considers to be 'political issues'.⁶⁵ Prior to 2020, the Electoral Commission (the UK's election regulator) used solely spending-based definitions to define 'political campaigning'. By August 2020, after consulting with social media platforms, devolved administrations and the Electoral Commission, the UK Government proposed that if digital content met both of the following two tests, then it would be considered as political campaigning content: (1) Where the material is intended to achieve the electoral success of registered political parties and candidates, or the material relates to a referendum; (2) Where digital content is promoted by either: (a)

⁶⁰ Murphy, L.W. 2020. *Facebook's Civil Rights Audit - Final Report*, 8 July. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

⁶¹ Bakir, V. 2010. *Sousveillance, Media and Strategic Political Communication: Iraq, USA, UK*. Bloomsbury.

⁶² Griffin, A. 2020. Twitter adds unprecedented warning to Trump tweet threatening to shoot Minneapolis protesters. *The Independent*, 29 May. <https://www.independent.co.uk/news/world/americas/twitter-trump-minneapolis-shoot-tweet-protesters-glorifying-violence-warning-george-floyd-a9538196.html>

⁶³ DCMS. 2018. *Disinformation and 'fake news': Interim Report*. 24 July. Digital, Culture, Media and Sport Committee, House of Commons 363, p.37. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/363/363.pdf>

⁶⁴ Facebook Business. 2020. *Sharing Our Actions on Stopping Hate*. 1 July. <https://www.facebook.com/business/news/sharing-actions-on-stopping-hate>

⁶⁵ Privacy International. 2019. https://privacyinternational.org/sites/default/files/2019-10/cop-2019_0.pdf

registered political parties, registered third party campaigners, candidates, holders of elected office and registered referendum campaigners - both paid and organic digital content; (b) other campaigners - paid digital content only.⁶⁶

Solutions to combat online hate speech and deceptive political ads. Stakeholders looking for solutions to combat online hate speech and deceptive political ads are divided on the value of microtargeting, but are more united on increasing the transparency of online political ads, and enabling advertisers to be held accountable on what they say. These solutions are examined below.

Solution: Divided Views on Whether Microtargeting for Political Ads should be Allowed

Facebook's tool, Dynamic Creative, allows campaigns to run thousands of different ads to enable optimisation for target audiences.⁶⁷ Stakeholders remain divided on the value of microtargeting for political ads. While some call for a total ban on microtargeted political ads, others (such as the UK Electoral Commission) note the value of a system of political advertising which preserves the ability to speak to (i.e. target) specific groups interested in a campaign's message.

The UK's Inquiry into Fake News and Disinformation concludes that it might be difficult to advocate a total ban on microtargeted political advertising online. Instead it recommended in July 2018 that there should be a ban on microtargeted political advertising to 'lookalikes' online (where people with similar views and opinions to core voters are also targeted), and that a minimum limit for the number of voters sent individual political messages should be agreed at a national level.⁶⁸ Limiting the number of ads that campaigns can run would increase the chances of them being held accountable for problematic messages, and would reduce the need for campaigns to collect large quantities of data about individuals for microtargeting purposes. This would bring ads back to a scale where the traditional accountability mechanism (journalism) can do its work.⁶⁹

Since 2017, the tech companies have gradually started to limit the level of detail that political campaigns or other groups can use to target voters.

- In October 2017, Twitter announced that it would update its policies for electioneering advertisers to limit targeting options.⁷⁰ Then, from 22 November 2019, Twitter said that globally, it would prohibit ads that contain references to political content including appeals for votes, solicitations of financial support, and advocacy for or against content that references a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive or judicial outcome.^{71 72} The reason Twitter gave for its ban was that it believed that the reach of political messages should be earned rather than bought. It states: 'A political message earns reach when people decide to follow an account or retweet. Paying for reach removes that decision, forcing highly optimized and targeted political messages on people. We believe this decision should not be compromised by money.'⁷³ However, Twitter still allows ads by non-political figures or organisations to advocate for political or social causes, on the grounds that cause-based advertising can facilitate public conversation around

⁶⁶ Cabinet Office & Smith, C. 2020. Press release: Political campaigning online to be made more transparent under new rules. 12 August.

<https://www.gov.uk/government/news/political-campaigning-online-to-be-made-more-transparent-under-new-rules>

⁶⁷ Bartlett, J., Smith, J. & Acton, R. 2018. *The Future of Political Campaigning*. Demos, p.33. <https://ico.org.uk/media/2259365/the-future-of-political-campaigning.pdf>

⁶⁸ DCMS. 2018. *Disinformation and 'fake news': Interim Report*. 24 July. Digital, Culture, Media and Sport Committee, House of Commons 363, p.38.

<https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/363/363.pdf>

⁶⁹ Who Targets Me. 2020. Ten simple ideas to regulate online political advertising in the UK. 4 June. <https://medium.com/@WhoTargetsMe/ten-simple-ideas-to-regulate-online-political-advertising-in-the-uk-52764b2df168>

⁷⁰ Falck, B. 2017. New Transparency for Ads on Twitter. *Twitter*, 24 October. <https://blog.twitter.com/official/enus/topics/product/2017/New-Transparency-For-Ads-on-Twitter.html>

⁷¹ There are exemptions. For instance, news publishers that meet Twitter's standards for certification as a news outlet are allowed to run political ads, but must not include advocacy for or against those topics.

⁷² Twitter. 2020. Political Content. <https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html>

⁷³ Collett, M. 2019. Twitter's Jack Dorsey announces ban on political advertising. *ABC News*, 30 October, <https://www.abc.net.au/news/2019-10-31/twitter-jack-dorsey-announces-ban-on-political-advertising/11656900>

important topics. These cause-based ads can address political issues such as the environment, economy or 'social equity' as long as they do not directly advocate for or against legislative, electoral, regulatory or judicial outcomes. Accompanying such ads are new restrictions on microtargeting. They can only be aimed at broad geographies (such as state, province, or region, but not postal codes) or non-political keywords or interests, and cannot target more personal characteristics such as race, gender or political affiliation.⁷⁴

- In November 2019, Google said that while it had 'never offered granular microtargeting of election ads', it was further limiting election adverts' targeting to the general categories of age, gender and general location (postal code level), as well as to contextual targeting;⁷⁵ that advertisers would no longer be able to target political messages based on users' interests inferred from their browsing or search histories; and that this approach would be enforced worldwide from January 2020.⁷⁶
- Facebook, however, continues to microtarget (with some exceptions for the US Facebook app).⁷⁷ It also still allows advertisers to use provocative targeting criterion, such as 'interested in "pseudoscience", thereby demonstrating that grouping users by their vulnerabilities continues.⁷⁸
- Apple's recent decision to increase privacy settings in iOS 14 may prove significant as it will require users to opt-in (rather than out) to sharing a device's IDFA (ID for Advertisers). Facebook stated, 'Our ability to deliver targeted ads on iOS 14 will be limited... as a result, some iOS 14 users may not see any ads from Audience Network, while others may still see ads from us, but they'll be less relevant.'⁷⁹

Some European context may also be useful. The behavioural advertising industry in the UK and Europe has been heavily criticised for how it is using personal data, especially around (1) transparency and consent, and (2) data supply chains. This had led the UK's Information Commissioner's Office (the national data regulator) to state that 'current practices remain problematic for the processing of personal data in general, even if the special category data were removed.'⁸⁰ Already highly significant, this regulatory interest is amplified because data used for elections relates to politics, and so constitutes sensitive personal data. The trade association, the Internet Advertising Bureau UK, have stated that advertising based on real time bidding (which makes use of behavioural data) can be conducted in a way that is compliant with European and UK legislation. It is, however, notable that this trade association is now actively advising advertisers about the merits of *contextual targeting*: this is a more traditional form of advertising that targets ads based on the content and environment that the ad will appear in, as opposed to data on the user that the ad is being served to.⁸¹

Solution: Increasing the transparency of online political ads

Diverse stakeholders agree on the need for maximum transparency for digital political advertising. At the time of the 2016 'Brexit' referendum (on the UK leaving the European Union) and the 2016

⁷⁴ Twitter. 2020. Cause-based advertising policy <https://business.twitter.com/en/help/ads-policies/ads-content-policies/cause-based-advertising.html>

⁷⁵ Spencer, S. 2019. An Update on Our Political Ads Policy. *Google*. <https://www.blog.google/technology/ads/update-our-political-ads-policy>

⁷⁶ Glazer, E. 2019. Facebook Weighs Steps to Curb Narrowly Targeted Political Ads. *The Wall Street Journal*. <https://www.wsj.com/articles/facebook-discussing-potential-changes-to-political-ad-policy-11574352887?mod=>

⁷⁷ By 2020 Facebook committed to a new advertising system for its US Facebook app so that advertisers running US housing, employment, and credit ads would no longer be allowed to target by age, gender, or postal code, and Facebook agreed to a much smaller set of targeting categories overall. (see Murphy, L. W. 2020. *Facebook's Civil Rights Audit - Final Report*, 8 July, pp.36-37. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf> p. 6).

⁷⁸ Angwin, J. 2020. Probing Facebook's Misinformation Machine. *The Markup*. <https://www.getrevue.co/profile/themarkup/issues/probing-facebook-s-misinformation-machine-241739>

⁷⁹ Facebook. 2020. Preparing Audience Network for iOS 14. <https://www.facebook.com/audiencenetwork/news-and-insights/preparing-audience-network-for-ios14/>

⁸⁰ Information Commissioner's Office. 2019. Update report into adtech and real time bidding, p.5-6. <https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf>

⁸¹ Internet Advertising Bureau UK. 2020. How to harness the power of contextual targeting. https://www.iabuk.com/sites/default/files/public_files/How-to-harness-the-power-of-contextual-targeting.%20pdf.pdf

US presidential election (events that caused widespread concern about the reach and impact of online disinformation) there were no ad libraries maintained by social media platforms. To address this, WhoTargetsMeWe, founded by Sam Jeffers and Louis Knight Webb in 2017, launched in the UK during the 2017 UK General Election to monitor the use of online political ads on Facebook in real time and provide analysis of their intended impact. By 2020, the Who Targets Me plug-in had been installed by over 30,000 users worldwide in more than 100 countries and was available in 25 languages.⁸² Since 2017 it has worked on over 10 elections worldwide including in Europe, the Middle East, and North and South America.⁸³

Since 2018, Google, Twitter and Facebook have provided publicly accessible, searchable libraries of election ads and spending on their platforms in the USA, with rollouts in select other countries.

- Google provides its ad library for election ads, namely: ads that feature a political party, a current elected officeholder, or a candidate in the European Union; election ads that feature or are run by a political party, a political candidate, or a current member of the lower house of the Indian Parliament; and ads that feature a current elected federal officeholder or candidate for the US House of Representatives, US Senate, President, or Vice President.⁸⁴
- Before November 2019 (when Twitter banned political ads), the transparency that Twitter provided depended on the ad and the country. Political ads tied to an election were archived for a longer period of time than political ads not tied to an election: the archive was available for Australia, Canada, European Union member states (limited to the 2019 Parliamentary elections), India and the USA. Commercial and political ads that are not tied to an election were archived for only seven days and with far more limited details on targeting and campaign performance.⁸⁵
- On Facebook, the ads are either classified as social issues, elections or politics or the advertisers self-declare that the ads require a 'Paid for by' disclaimer; and users are able to click on political, electoral, or social issue ads to access information about the ad's reach, who was shown the ad, and the entity responsible. Facebook has since taken more measures to increase the transparency of Political Ads and its Public Ad Library. By 2020, this included the following for the US Facebook app.
 - Users can see information about why they received the ad, namely which of the ad targeting categories selected by the advertiser the user fell into. In addition, the same pop-up that appears by clicking on an ad gives users more control over the ads they see.
 - Facebook updated its Ad Library to increase transparency and provide useful data to the public. This was to address the civil rights community's concerns about the Ad Library's search functionality and so that the Ad Library could be a more effective tool for identifying patterns of political suppression and misinformation.⁸⁶
 - Ensuring that ads about social issues, elections or politics indicate the name of the person or entity responsible for the ad through a disclaimer displayed when the ad is shown, and that the 'sponsored' designation and the disclaimer designation is retained when the ad is shared or forwarded.

Unfortunately, social media companies have yet to roll out such libraries to many countries

⁸² English, Spanish, German, Italian, Hungarian, Russian, Latvian, Portuguese, Swedish, Finnish, French, Dutch, Croatian, Ukrainian, Hebrew, Bulgarian, Czech, Estonian, Greek, Macedonian, Polish, Romanian, Slovakian, Slovenian and Maltese.

⁸³ WhoTargetsMe 2020. <https://chrome.google.com/webstore/detail/who-targets-me/fcejbjalmgocomoinikjejnkimlnoljp?hl=en>

⁸⁴ Privacy International. 2019. https://privacyinternational.org/sites/default/files/2019-10/cop-2019_0.pdf

⁸⁵ Twitter, 2020. <https://business.twitter.com/en/help/ads-policies/ads-transparency-center-faqs.html>

⁸⁶ Murphy, L. W. 2020. *Facebook's Civil Rights Audit - Final Report*, 8 July, pp.36-37. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

beyond the USA. A recent study of cybersecurity in elections in Commonwealth countries recommends that: 'Commonwealth countries should consider legislating to ensure that platforms and advertising networks are obliged to make political adverts public, in line with best practices in the area which allow public research and scrutiny'.⁸⁷

Furthermore, in 2019, Privacy International records that the targeting information provided by Facebook, Google, and Twitter is inadequate as it is still impossible to meaningfully understand who political advertisers are targeting across the three platforms. The ad libraries of Facebook, Google, and Twitter, in varying degrees, provide broad ranges of targeting information on some ads in some countries, instead of meaningful insight into how an ad was targeted. This is especially problematic given the granularity with which advertisers, political or not, are able to microtarget users. Google is especially deficient given that it only provides broad ranges of insight, such as 100K-1M as the number of times an ad was shown, rather than meaningful information about how an ad or campaign was targeted.⁸⁸

Solution: Holding advertisers accountable for what they say

Various solutions have been proposed to enable advertisers to be held accountable for what they say and for breaking rules. These include the following:

- *Force strong verification.* Companies accepting political ads should publish the name and address of the ultimate entity paying for ads and feel confident that the source of their funds is clear.⁸⁹ Similarly, tech companies should address the issue of shell corporations and other professional attempts to hide identity in ad purchasing, especially around election advertising.⁹⁰
- *Make clear what is an ad.* Social media advertising blurs the lines between paid and organic content. Creating inflammatory ads for the purpose of sharing to attract additional 'free' views means that the content of political ads' is being led by the incentives of the platform, rather than the value of the arguments they make. Paid political material should be permanently labelled as an ad.⁹¹
- *Ensure these measures are 'always on'* as political actors seek to shape debates long before election campaigns start.⁹²

Addressing these recommendations, in August 2020, the UK government published proposals calling for *digital imprints* (to clearly show who the campaigners are) to apply to all types of campaign content regardless of the country it is being promoted from, across a variety of digital platforms, with the regime intended to apply all year round.⁹³

Other recommendations by civil society have yet to be embraced by the UK government. For instance, one such recommendation is to *introduce a blackout period for political advertising*. Social media ads launched on the day of an election can introduce new, untested and potentially influential claims into a campaign at the precise moment voters are making their final decision. For instance, in the 2016 Brexit referendum, having identified from focus groups that crucial swing voters were confused, and liable to change their decision on which way to vote based on whether they had last seen a message from either side of the referendum campaign, Vote Leave decided to ensure that a Vote Leave ad was delivered to swing voters as late as possible in

⁸⁷ Brown, I., Marsden, C.T., Lee, J., & Veale, M. 2020. *Cybersecurity for elections. A Commonwealth guide on best practice*. P.12.

<https://thecommonwealth.org/sites/default/files/inline/Commonwealth%20cybersecurity%20for%20elections%20guide.pdf>

⁸⁸ Privacy International. 2019. https://privacyinternational.org/sites/default/files/2019-10/cop-2019_0.pdf

⁸⁹ Who Targets Me. 2020. Ten simple ideas to regulate online political advertising in the UK. 4 June. <https://medium.com/@WhoTargetsMe/ten-simple-ideas-to-regulate-online-political-advertising-in-the-uk-52764b2df168>

⁹⁰ DCMS. 2018. *Disinformation and 'fake news': Interim Report*. 24 July. Digital, Culture, Media and Sport Committee, House of Commons 363, p.38. <https://publications.parliament.uk/pa/cm201719/cmselect/cmcmds/363/363.pdf>

⁹¹ Who Targets Me. 2020. Ten simple ideas to regulate online political advertising in the UK. 4 June. <https://medium.com/@WhoTargetsMe/ten-simple-ideas-to-regulate-online-political-advertising-in-the-uk-52764b2df168>

⁹² Who Targets Me. 2020. Ten simple ideas to regulate online political advertising in the UK. 4 June. <https://medium.com/@WhoTargetsMe/ten-simple-ideas-to-regulate-online-political-advertising-in-the-uk-52764b2df168>

⁹³ Cabinet Office & Smith, C. 2020. Press release: Political campaigning online to be made more transparent under new rules. 12 August. <https://www.gov.uk/government/news/political-campaigning-online-to-be-made-more-transparent-under-new-rules>

the campaign.⁹⁴ With other sources of media having their own blackout on election day, voters have little chance of establishing whether the claims they see are true. All political advertising, other than ‘get out the vote’ advertising of the format ‘Vote X’, should therefore be banned in the 24 hours before voters go to the polls.⁹⁵

Enforcement is a key element of holding political campaigners and advertisers to account, and Facebook has been severely criticised on these grounds, despite its new policies that prohibit ads targeting the US with messages discouraging people from voting or designed to suppress participation in the census. For instance, the civil rights audit of Facebook’s US app notes that Facebook’s decisions about posts from President Donald Trump related to mail-in-ballots in Michigan, Nevada and California in May 2020 will encourage others to use the platform to spread damaging misinformation about voting.⁹⁶ In May 2020, Trump made a series of posts that labelled official, state-issued ballots or ballot applications ‘illegal’ and gave false information about how to obtain a ballot, such as the following.⁹⁷

‘There is NO WAY (ZERO!) that Mail-In Ballots will be anything less than substantially fraudulent. Mail boxes will be robbed, ballots will be forged & even illegally printed out & fraudulently signed. The Governor of California is sending Ballots to millions of people, anyone living in the state, no matter who they are or how they got there, will get one. That will be followed up with professionals telling all of these people, many of whom have never even thought of voting before, how, and for whom, to vote. This will be a Rigged Election. No way!’⁹⁸

Facebook’s voter interference policy prohibits false misrepresentations regarding the ‘methods for voting or voter registration’ and ‘what information and/or materials must be provided in order to vote.’ In California, ballots were not being issued to ‘anyone living in the state, no matter who they are’. In fact, in order to obtain a mail-in ballot in California, one has to register to vote. Despite these facts, Facebook decided that these posts did not violate its policies, a stance that elicited uproar from civil rights leaders, elected officials and Facebook’s staff.⁹⁹

⁹⁴ Cummings, D. 2017. On the referendum #22: Some basic numbers for the Vote Leave campaign. <https://dominicummings.com/2017/01/30/on-the-referendum-22-some-numbers-for-the-vote-leave-campaign/>. Howard, P.N. 2018. Claim No: CO/3214/2018. THE QUEEN on the application of SUSAN WILSON & OTHERS -and-THE PRIME MINISTER, Report of Dr Philip N. Howard Professor, Oxford University to the High Court of Justice, Queen’s Bench Division, Administrative Court. <https://www.ukineuchallenge.com/wp-content/uploads/2018/12/257136-Expert-report-of-Prof-Howard-FINAL-Signed.pdf>

⁹⁵ Who Targets Me. 2020. Ten simple ideas to regulate online political advertising in the UK. 4 June. <https://medium.com/@WhoTargetsMe/ten-simple-ideas-to-regulate-online-political-advertising-in-the-uk-52764b2df168>

⁹⁶ Murphy, L.W. 2020. *Facebook’s Civil Rights Audit – Final Report*, 8 July. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

⁹⁷ Murphy, L.W. 2020. *Facebook’s Civil Rights Audit – Final Report*, 8 July, pp.37-38. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

⁹⁸ Trump, D.J. 2020. Facebook. 26 May. <https://www.facebook.com/DonaldTrump/posts/10164748538560725>

⁹⁹ Murphy, L.W. 2020. *Facebook’s Civil Rights Audit – Final Report*, 8 July, pp.37-38. <https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf>

Quick, Uncontroversial Fixes

Many actions have been taken by social media companies, and some governments to address the advertising-oriented electoral harms that arise from social media. We recommend that governments should take the following two solutions as relatively quick and uncontroversial fixes (notwithstanding other issues that this submission does not address such as a detailed assessment of the legality of behavioural targeting, digital literacy, foreign influence campaigns and the declining health of the wider news media landscape).

1. **Increase the transparency of online political ads by ensuring that social media platforms that allow political ads and political issue ads keep accessible, searchable libraries of these ads.** As Privacy International observed in September 2019, these companies have applied different policies in different countries. In jurisdictions where these companies are under pressure to act (by governments, institutions such as the European Union, or civil society), they have adopted self-regulatory practices. Where such pressure is absent, they have largely failed to act. Most users around the world therefore lack meaningful insight into how ads are being targeted through these platforms.
 - a. **Facebook** provides heightened transparency for political ads (i.e. where political advertisers become authorised, political ads carry disclosures, and ads are archived) in only 35 countries (roughly 17% of the countries in the world). Political issue ads are enforced in even fewer countries. For either political ads or political issue ads, Australia does not appear to be one of these countries.
 - b. **Google** provides heightened transparency for political ads in 30 countries (roughly 15% of the countries in the world). Australia does not appear to be one of these countries.
 - c. Until late 2019, **Twitter** provided heightened transparency for ads tied to specific elections (rather than political ads more generally), but in only 32 countries (roughly 16% of the countries in the world).¹⁰⁰ In Australia, this applied to political ads but not to political issue ads, and only to certain elections. Since November 2019, Twitter has banned political ads globally, but still allows cause-based ads that can address political issues such as the environment, economy or social equity as long as they do not directly advocate for or against legislative, electoral, regulatory or judicial outcomes.
2. **Ensure that there are digital imprints to hold advertisers accountable for what they say.** Ensure that they apply to all types of campaign content regardless of the country it is being promoted from, across all digital platforms, with the regime applied all year round.

Longer-term Investigation

1. **Close and sincere attention should be paid to the legality of microtargeting.** As well as the novel question of potential contravention of the right to freedom of thought, UK and European developments are challenging the questionable legal premises of microtargeting (on the basis of consent, transparency and data supply chains). Australian and Victorian regulators (data protection, media and electoral) should reflect on whether the law is being truly and meaningfully applied to behavioral advertising and microtargeting.
2. **More user-based and legal studies are needed into the impact of microtargeting on electorates,** both in terms of direct influence (including the nature and extent of undue influence) and regarding wider trust in the fair conduct of elections.

¹⁰⁰ Privacy International. 2019. https://privacyinternational.org/sites/default/files/2019-10/cop-2019_0.pdf