# DISINFORMATION, DEEPFAKES AND DEMOCRACIES: THE NEED FOR LEGISLATIVE REFORM

ANDREW RAY*

*Rapid technological advancement is changing the way that political parties, voters, and media platforms engage with each other. This along with cultural change has led to an emerging era of disinformation and misinformation driven by both domestic and foreign actors. Political deepfakes, videos created through the use of artificial intelligence, allow individuals to rapidly create fake videos indistinguishable from true content. These videos have the capacity to undermine voter trust and could alter electoral outcomes. Regulating disinformation however raises significant free speech concerns, as well as questions about where liability should fall. In particular, holding large technology and media platforms accountable for content could lead to unintended chilling effects around freedom of expression, harming rather than protecting democratic institutions. Proposed regulations should therefore be carefully analysed through the framework of the implied freedom of political communication, ensuring that any new laws are proportionate and tailored to the threat they seek to prevent. This article analyses how current Australian law interacts with political deepfakes and proposes two targeted amendments to our federal electoral regulations to reduce the threat they pose to elections.*

## I INTRODUCTION

The rapid advancement of artificial intelligence ('AI') and machine-learning algorithms ('MLAs') is disrupting the way that we operate and do business.[1] The

* BSc/LLB (Hons I) (ANU), Visiting Fellow at Australian National University College of Law. The author would like to thank Associate Professor Philippa Ryan and the anonymous reviewers and editors whose comments strengthened this article. This article reflects the author's personal views.

1 While much of the underpinning logic behind AI and MLAs has been understood since the 1970s, it is the rapid advancement in computing power, combined with increasing data gathering and analysis capabilities that is driving the growth in AI: see Andrea Zanella et al, 'Internet of Things for Smart Cities' (2014) 1(1) *Internet of Things Journal* 22; Monika Zalnieriute, Lyria Bennett Moses and George Williams, 'The Rule of Law and Automation of Government Decision-Making' (2019) 82(3) *Modern Law Review* 425; Will Bateman, 'Algorithmic Decision-Making and Legality: Public Law Dimensions' (2020) 94(7) *Australian Law Journal* 520. Given the rapidly moving field of technology law (and deepfake technology in particular), this article draws on grey literature to supplement peer-reviewed research. For discussion on grey literature in the context of evolving medical technology, see Louisa Degenhardt et al, 'Searching the Grey Literature

interaction between AI and law, and the day-to-day operation of government are posing unique challenges, given the speed at which AI operates and the threat it presents to accountability and transparency of government. This has been demonstrated in an Australian context through the challenges driven by automated decision-making,[2] including the ongoing Robodebt debacle.[3] While much has been written about the application of AI to government,[4] little analysis has been conducted regarding the threat AI poses to elections, and by extension to the foundations of representative democracies. In turn, this means few protections are available to combat this threat.

This article outlines the limitations of existing law as applied to the emerging problem of 'political deepfakes', a subtype of political disinformation. Deepfakes are videos created using AI, which allow creators to superimpose images and audio from one video to another.[5] In effect, deepfake technology allows a user to create a fake video of a person saying or doing almost anything, only limited by their creativity and the footage of the subject they can source. Regulating deepfakes poses unique challenges in an Australian context through the operation of the implied freedom of political communication. Similarly, there remain significant challenges when designing regulations to ensure that speech is not overburdened and that regulations are proportionate and tailored to the threat they seek to prevent.

This article proceeds in four parts. Part II analyses the threat posed to Australian elections by political deepfakes. Parts III and IV explore current private and public remedies available to legitimate political actors and the Australian Electoral Commission ('AEC') to combat political deepfakes. The insufficiency of these available remedies to mitigate the harms caused by political deepfakes is then examined. Part V proposes legislative reform via a model law that could be enacted by the Commonwealth, state and territory governments to combat political deepfakes. In doing so, the article recommends against broader regulation of misinformation and disinformation which may lead to a significant chilling effect on political communication.

---

to Access Research on Illicit Drug Use, HIV and Viral Hepatitis' (Technical Report No 334, National Drug and Alcohol Research Centre, University of New South Wales, 2016).

2 Andrew Ray, 'Implications of the Future Use of Machine Learning in Complex Government Decision-Making in Australia' (2020) 1(1) *Australian National University Journal of Law and Technology* 4.

3 Richard Glenn, Acting Commonwealth Ombudsman, 'Centrelink's Automated Debt Raising and Recovery System' (Report No 2, April 2017) 7–8 [3.2]–[3.6] <https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf>; Order of Davies J in *Amato v Commonwealth* (Federal Court of Australia, VID611/2019, 27 November 2019). The settlement was approved by the Federal Court in *Prygodicz v Commonwealth [No 2]* [2021] FCA 634; however, accountability issues remain as the opposition pushes for review of the decisions leading to the class action.

4 See, eg, Zalnieriute, Bennett Moses and Williams (n 1).

5 Kristina Libby, 'Deepfakes Are Amazing. They're Also Terrifying for Our Future', *Popular Mechanics* (online, 13 August 2020) <https://www.popularmechanics.com/technology/security/a28691128/deepfake-technology/>.

## II DEEPFAKES AND DEMOCRACIES

In the context of elections, AI combined with key datasets (commonly referred to as Big Data) is being used by political parties to better target swing voters and to assess the palatability of policy positions.[6] Similarly, electoral agencies are using algorithms to manage the increasingly complex process of counting votes.[7] These algorithms are not subject to public scrutiny.[8] While these issues are concerning, the threats they pose can largely be mitigated through open, fair and transparent electoral processes. This is because electoral agencies are responsible to Parliament, and therefore the population can decide whether the actions of political parties (and the AEC) should be punished at the ballot box.[9] It is therefore the influence of AI on the conduct and results (rather than the management) of elections that is the primary focus of this article.

### A Political Deepfakes

The use of AI technologies represents a significant and growing threat to electoral security. In particular, deepfake technology when deployed by experts can create videos of politicians so realistic they cannot be distinguished from a real video by humans or computers designed to detect them.[10] Deepfakes are created using 'neural networks that analyze large sets of data … to learn to mimic a person's facial expressions, mannerisms, voice, and inflections'.[11] By way of a popular example, similar technology was used to create scenes in which the late Carrie Fisher appeared in the recent Star Wars film: *Rogue One*.[12]

Historically, individuals wishing to make a useful (or, perhaps more accurately described, *undetectable*) deepfake, required hundreds of images of their 'subject' to train an MLA.[13] However, recent advances in technology have meant that only

---

6 Jennifer Lees-Marshment et al, 'Vote Compass in the 2014 New Zealand Election' (2015) 67(2) *Political Science* 94.

7 Ben Raue, 'Looking Out for No 1: Why the Senate Vote Count Needs Greater Transparency', *The Guardian* (online, 20 July 2016) <https://www.theguardian.com/australia-news/2016/jul/20/looking-out-for-no-1-why-the-senate-vote-count-needs-greater-transparency>.

8 *Cordover and Australian Electoral Commission (Freedom of information)* [2015] AATA 956 (11 December 2015); Ray (n 2) 13–14.

9 Brian Galligan, 'Parliamentary Responsible Government and the Protection of Rights' (Papers on Parliament No 18, Parliament of Australia, December 1992).

10 Mika Westerlund, 'The Emergence of Deepfake Technology: A Review' (2019) 9(11) *Technology Innovation Management Review* 39, 45–6.

11 Ibid 40.

12 Erin Winick, 'How Acting as Carrie Fisher's Puppet Made a Career for Rogue One's Princess Leia', *MIT Technology Review* (online, 16 October 2018) <https://www.technologyreview.com/2018/10/16/139739/how-acting-as-carrie-fishers-puppet-made-a-career-for-rogue-ones-princess-leia/>. In an Australian context, fans have inserted the Joker into *A Knight's Tale* (Columbia Pictures, 2001): Ben Gilbert, 'An Incredible Series of Videos Swap Famous Hollywood Faces to Demonstrate How Convincing "Deepfake" Tech Has Gotten: Take a Look', *Business Insider Australia* (online, 31 May 2019) <https://www.businessinsider.com.au/deepfakes-of-famous-movies-youtube-channel-2019-5?r=US&IR=T>.

13 See, eg, Supasorn Suwajanakorn, Steven M Seitz and Ira Kemelmacher-Shlizerman, 'Synthesizing Obama: Learning Lip Sync from Audio' (2017) 36(4) *ACM Transactions on Graphics* 1.

a small number of images are required to generate realistic videos of the subject.[14] This, combined with the fact that videos shot front-on in consistent light are the easiest to replicate,[15] makes political figures a ripe target for deepfakes. This is due to the wide availability of footage of political figures in which they are positioned forward-facing, under similar lighting conditions.[16] This ease of creation is demonstrated by the fact that deepfakes can now be created on a smartphone, using only a few images of the intended subject.[17]

The targeting of politicians with deepfake technology is more than an academic hypothesis. Indeed, deepfakes have been made featuring Donald Trump,[18] Barack Obama,[19] Manoj Tiwari,[20] Vladimir Putin[21] and Sophie Wilmès.[22] These examples, while well-known, are not exhaustive. The targeting of then Belgian Prime Minister Sophie Wilmès by Extinction Rebellion[23] in mid-2020 is of particular concern as it appears to be the *first* adverse targeting of a politician: previous examples of political deepfakes were generally educational, comedic or satirical.[24] The video in question, which showed Wilmès giving a fictitious speech about the link between COVID-19 and climate change, was widely shared on social media. Critically, at least some users were tricked into believing the video was real.[25] Regardless of whether you agree with the motivation behind the video, the use of deepfake technology to falsely attribute a speech to an elected Prime Minister is of grave concern.

---

14 Egor Zakharov et al, 'Few-Shot Adversarial Learning of Realistic Neural Talking Head Models', *arXiv* (submitted 20 May 2019, revised 25 September 2019) <https://arxiv.org/abs/1905.08233>.

15 'How to Create the Perfect DeepFakes', *Alan Zucconi* (Blog Post, 14 March 2018) <https://www.alanzucconi.com/2018/03/14/create-perfect-deepfakes/>.

16 For example, politicians regularly appear at press conferences and in news segments where they are often filmed looking directly at the camera in a well lit environment.

17 See, eg, NEOCORTEXT, INC., 'Reface: Face Swap Videos', *Apple App Store* (Application, 2020) <https://apps.apple.com/app/id1488782587>.

18 Helena Skinner, 'French Charity Publishes Deepfake of Trump Saying "AIDS is over"', *Euronews* (online, 9 October 2019) <https://www.euronews.com/2019/10/09/french-charity-publishes-deepfake-of-trump-saying-aids-is-over>.

19 James Vincent, 'Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA about Fake News', *The Verge* (online, 17 April 2018) <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed>.

20 Regina Mihindukulasuriya, 'Why the Manoj Tiwari Deepfakes Should Have India Deeply Worried', *The Print* (online, 29 February 2020) <https://theprint.in/tech/why-the-manoj-tiwari-deepfakes-should-have-india-deeply-worried/372389/>. This video differs from the other examples as it was made by the subject to help them communicate to voters with different language backgrounds.

21 Karen Hao, 'Deepfake Putin Is Here to Warn Americans about Their Self-Inflicted Doom', *MIT Technology Review* (online, 29 September 2020) <https://www.technologyreview.com/2020/09/29/1009098/ai-deepfake-putin-kim-jong-un-us-election/>.

22 'The Truth about COVID-19 and the Ecological Crisis: A Speech for Sophie Wilmès', *Extinction Rebellion Belgium* (Web Page, April 2020) <https://www.extinctionrebellion.be/en/tell-the-truth>.

23 Ibid.

24 Westerlund (n 10) 43.

25 Gerald Holubowicz, 'Extinction Rebellion S'empare des Deepfakes en Belgique' [Extinction Rebellion Takes over Deepfakes in Belgium], *Mediapart* (Blog Post, 15 April 2020) <https://blogs.mediapart.fr/geraldholubowicz/blog/150420/extinction-rebellion-s-empare-des-deepfakes-en-belgique>.

## B Impact on Elections

This article will focus on two primary threats posed to elections by deepfakes: the use of deepfakes to alter voter preferences, and the impact of deepfakes on trust generally in elections and democratic institutions.[26] First, through their potential impact on voter preferences, deepfakes may be used to obfuscate or undermine a politician's (or political party's) stance on a given issue, or to target their credibility. Given the shift to longer periods of pre-polling in Australia (and other democracies),[27] the release of a deepfake within this period or just before election day will make it extremely challenging for politicians to respond before any votes are cast. For example, a deepfake of a politician with a strong anti-drug platform consuming an illicit drug could be both impactful, and difficult to disprove.[28] A deepfake could be made as part of a candidate's official campaign, by an overseas actor attempting to sway an election, or even by an individual disconnected from the political process.

While there is no evidence that deepfakes have impacted an Australian election to date, compromising (albeit true) video footage has previously led to federal candidates dropping out of an electoral race.[29] Meanwhile, doctored footage has been used in the United States ('US') by the Republican Party to attack House Speaker Nancy Pelosi by slowing down real video clips of her speeches to slur her words and make her appear drunk.[30] Similar videos were also used to target President Joe Biden in the 2020 Presidential election, with experts warning prior to the election that the worst was yet to come as 'cutting-edge methods such as deepfakes are best suited to … predictable moment[s] of public uncertainty'.[31] Such a moment, they posited, would occur following the election, with Trump hinting

26 Secondary threats could include undermining diplomacy and jeopardising national security. These threats can be viewed as subsidiary to the primary threats identified above in that they rely on either convincing a particular actor a fake video is real or in eroding public trust in video content, for example, fake news about nuclear attacks could cause general panic and reduce trust in future warnings.

27 Stephen Mills and Martin Drum, 'Surge in Pre-poll Numbers at 2019 Federal Election Changes the Relationship between Voters and Parties', *The Conversation* (online, 19 August 2019) <https://theconversation.com/surge-in-pre-poll-numbers-at-2019-federal-election-changes-the-relationship-between-voters-and-parties-121929>. This trend has increased in recent elections: Damon Muller, 'Trends in Early Voting in Federal Elections', *Parliament of Australia* (Web Page, 8 May 2019) <https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/FlagPost/2019/May/Trends_in_early_voting_in_federal_elections>.

28 Further possibilities could include footage of candidates withdrawing from a race and endorsing another candidate, a politician committing an offence, accepting a bribe, or outlining a fake policy position. Given the ease of use of the technology, users are limited only by their creativity.

29 Josh Bavas, 'One Nation Election Candidate Steve Dickson Resigns over Strip Club Videos', *ABC News* (online, 30 April 2019) <https://www.abc.net.au/news/2019-04-30/one-nation-candidate-steve-dickson-quits-over-strip-club-video/11056676>.

30 Hannah Denham, 'Another Fake Video of Pelosi Goes Viral on Facebook', *The Washington Post* (online, 3 August 2020) <https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/>.

31 Clint Watts and Tim Hwang, 'Deepfakes Are Coming for American Democracy: Here's How We Can Prepare', *The Washington Post* (online, 10 September 2020) <https://www.washingtonpost.com/opinions/2020/09/10/deepfakes-are-coming-american-democracy-heres-how-we-can-prepare/>.

that he would not accept electoral defeat.[32] That set of circumstances unfolded partly as predicted with Trump declaring the election results 'fake news' and his supporters storming the Capitol in circumstances condemned as terrorism by US security agencies.[33] There was however no detectable use of deepfake videos, with the potential for a faked video of then President-elect Biden accepting 'defeat' remaining only a possibility. It is noteworthy that despite public institutions, inquiries and courts all labelling the fraud claims false, Trump and the Republican Party more broadly continue to push the electoral fraud claims publicly.

### *1 Changing Voter Preferences*

Exactly how many voters could be misled by a deepfake remains unclear. However, if marginal seats were targeted during an election, even swaying as few as 100 voters could be impactful.[34] In this context, a 2020 study found that approximately 15% of viewers in a controlled trial believed a deepfake of Obama was real.[35] While it is unlikely that everyone who believes a deepfake will alter their vote because of it (in part due to the strength of party allegiance),[36] the possibility should not be discounted. Indeed, it may not be necessary for voters to alter their vote for a deepfake video to impact an election. For example, deepfake videos could force candidates to withdraw or impact a candidate's or party's fundraising ability – these results themselves having an indirect effect on electoral outcomes. Further, while some authors have found that disinformation generally has little direct impact on elections,[37] disinformation has been shown to have (at least some) impact in Australian elections. For example, the Australian Labor Party acknowledged the impact of the (false) 'death tax' ads on its 2019 campaign, although they accepted that this alone did not decide the election.[38] Additionally, while disinformation (and specifically, in the context of this article, the use of deepfakes) may not alter which party secures a majority of seats, it may play a larger role in deciding *individual* electoral contests. This is especially the case with deepfakes, where, as discussed

32 'Donald Trump Refuses to Commit to Peaceful Transfer of Power if He Loses US Election', *ABC News* (online, 24 September 2020) <https://www.abc.net.au/news/2020-09-24/donald-trump-wont-commit-to-transfer-of-power-after-election/12696786>.

33 See generally 'FBI Chief Calls Capitol Attack Domestic Terrorism and Rejects Trump's Fraud Claims', *The Guardian* (online, 11 June 2021) <https://www.theguardian.com/us-news/2021/jun/10/capitol-attack-fbi-christopher-wray-congress>.

34 For example, in the 2020 Northern Territory election 11/25 seats would have changed hands if 100 voters had been swayed by a deepfake: 'NT Summary of Two Candidate Preferred Votes by Division', *Northern Territory Electoral Commission* (Web Page, 2020) <https://ntec.nt.gov.au/elections/2020-territory-election/results/nt-summary-of-two-candidate-preferred-votes-by-division>. The average turnout for each division was 4,235 voters, so swaying ~2.5% of voters could have altered 11/25 contests.

35 Cristian Vaccari and Andrew Chadwick, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News' (2020) 6(1) *Social Media + Society* 1, 6.

36 Spencer McKay and Chris Tenove, 'Disinformation as a Threat to Deliberative Democracy' (2020) (July) *Political Research Quarterly* 1, 1.

37 Ibid. However, the authors went on to assess other harms that disinformation may pose, including degrading trust in media organisations and academic think tanks.

38 See, eg, Craig Emerson and Jay Weatherill, 'Review of Labor's 2019 Federal Election Campaign' (Report, 7 November 2019) 79–80.

above, it is possible for actors to target individual politicians by, for example, creating a deepfake of them engaging in illegal conduct. In this context, critically, at a federal level Australia remains vulnerable to targeted attacks: 36 lower house seats are currently held by a margin of less than 5%, 84 by less than 10% and 129 by less than 15%.[39]

It is however the secondary threat that is likely of greater concern. In addition to the percentage who believed the deepfake was real, the 2020 study found that only 50.8% of the participants were *not* deceived by the video.[40] The remainder were *unable* to determine if the video was real or fake. It is this segment of individuals that highlights the second threat posed by deepfakes to elections: a reduction in trust in video footage and news impacting our perception of democracy more broadly.

### *2 Decreasing Trust in Democracy and Democratic Institutions*

Increasingly, Australians are turning to digital platforms such as Facebook to access news content.[41] This mirrors a global trend towards accessible and shareable content,[42] which is making it easier for fake news to be distributed widely. The shift to digital content has coincided with decreasing trust in politicians and politics in general.[43] Political deepfakes will further erode trust by allowing candidates to deride real footage as fake news, feeding into increasing claims by politicians that they have been set up.[44] It is this threat that most alarms political scientists as, after all, threats to a single election are of themselves a threat to democracy.[45] However, the rise of disinformation more broadly has the capacity to fundamentally undermine 'truth' in elections with disastrous consequences. For example, in the US, disproven rumours of electoral fraud are supporting a wave of electoral reforms that will make it harder to vote to 'safeguard' future elections.[46] These laws

---

39 Corresponding to 24%, 56% and 85% of lower house seats accordingly. Analysis conducted on AEC data from the recent 2019 federal election and 2020 Eden-Monaro by-election: Australian Electoral Commission, 'Seat Summary', *Tally Room 2019 Federal Election* (Web Page, 2019) <https://results.aec.gov.au/24310/Website/HouseSeatSummary-24310.htm> (results on file with author).

40 This was described as 'surprising given the statement [an unsophisticated insult about Donald Trump] was highly improbable': Vaccari and Chadwick (n 35) 6.

41 Australian Competition and Consumer Commission, 'Digital Platforms Inquiry' (Final Report, June 2019) ch 1; See also Christopher Hughes, 'News Sources in Australia in 2021', *Statista* (online, 12 July 2021) <https://www.statista.com/statistics/588441/australia-news-sources/>.

42 Katie Elson Anderson, 'Getting Acquainted with Social Networks and Apps: Combating Fake News on Social Media' (2018) 35(3) *Library Hi Tech News* 1.

43 Simon Tormey, 'The Contemporary Crisis of Representative Democracy' (Papers on Parliament No 66, Parliament of Australia, October 2016) 90 <https://www.aph.gov.au/About_Parliament/Senate/Powers_practice_n_procedures/pops/Papers_on_Parliament_66/The_Contemporary_Crisis_of_Representative_Democracy>; Russell J Dalton, *Democratic Challenges, Democratic Choices: The Erosion of Political Support in Advanced Industrial Democracies* (Oxford University Press, 2004).

44 See, eg, comments made by then President Donald Trump during the 2020 election: David Smith, 'Wounded by Media Scrutiny, Trump Turned a Briefing into a Presidential Tantrum', *The Guardian* (online, 14 April 2020) <https://www.theguardian.com/us-news/2020/apr/13/trump-coronavirus-meltdown-media-authority>.

45 McKay and Tenove (n 36).

46 Sam Levine, 'The Republicans' Staggering Effort to Attack Voting Rights in Biden's First 100 Days', *The Guardian* (online, 28 April 2021) <https://www.theguardian.com/us-news/2021/apr/28/republicans-voter-suppression-biden-100-days>.

have been held constitutional by the US Supreme Court,[47] and may, along with gerrymandering, decide the outcome of future elections alone notwithstanding for whom people vote on voting day. Deepfakes may exacerbate these underlying issues and cause distrust amongst voters themselves who may not know *whom* or *what* they can actually trust, allowing lawmakers to pass anti-democratic laws to 'safeguard' elections.

These threats are not insignificant, especially as deepfakes can be generated and shared from within or outside of Australia by anyone with a desktop computer or smartphone.[48] It is this accessibility that makes the threat most concerning, as once the videos have been created and shared, they can be re-uploaded rapidly making it almost impossible for them to be taken down (even if proven false). For example, the widely discredited video *Plandemic* was repeatedly re-uploaded to alternative hosting sites after being taken down by Facebook and YouTube, with commentators suggesting the attempt to shut down the video led to it being viewed by a wider audience.[49]

### C Increasing Challenge of Electoral Interference

The threat posed by deepfakes is heightened by the increasing level of foreign interference in elections. The threat posed by foreign actors is unique, in that they can operate outside a target jurisdiction, while still being able to spread fake news through social media. This rise in foreign interference both increases the likelihood that deepfakes will be used and makes them harder to combat due to limitations of domestic law. Despite these limitations, difficulties in attributing disinformation to a state mean that domestic regulations are likely more useful than pursuing action internationally.[50]

Foreign interference impacted the outcome of the 2016 US Presidential election,[51] and has been of increasing concern to the Australian Government. For example, the Government has recently launched Senate inquiries into foreign interference,[52] proposed a widening of the Australian Security Intelligence Organisation's powers

47 *Brnovich v Democratic National Committee*, 594 US ___ (2021). For commentary: see, eg, Lauren Fedor, 'US Supreme Court Upholds Arizona Law in Voting Rights Challenge', *Financial Times* (online, 2 July 2021) <https://www.ft.com/content/35e67872-e1eb-449d-8745-3d0c13db1526>.

48 Best results require a mid-high end graphics card: Timothy B Lee, 'I Created My Own Deepfake: It Took Two Weeks and Cost $552', *ARS Technica* (online, 16 December 2019) <https://arstechnica.com/science/2019/12/how-i-created-a-deepfake-of-mark-zuckerberg-and-star-treks-data/>.

49 Andrea Bellemare, Katie Nicholson and Jason Ho, 'How a Debunked COVID-19 Video Kept Spreading after Facebook and YouTube Took It Down', *CBC News* (online, 21 May 2020) <https://www.cbc.ca/news/technology/alt-tech-platforms-resurface-plandemic-1.5577013>.

50 Björnstjern Baade, 'Fake News and International Law' (2019) 29(4) *European Journal of International Law* 1357, 1361–2. This article will therefore focus on domestic rather than international law.

51 United States Senate Select Committee on Intelligence, *Russian Active Measures Campaigns and Interference in the 2016 US Election* (Report, 2020) vol 5 <https://www.intelligence.senate.gov/sites/default/files/documents/report_volume5.pdf>; 'Russia Worked to Help Trump in 2016 Election: Senate Panel', *Aljazeera* (online, 18 August 2020) <https://www.aljazeera.com/news/2020/8/18/russia-worked-to-help-trump-in-2016-election-senate-panel>. The US federal government has implemented laws encouraging research deepfakes but is yet to legislate to directly combat the threat: *National Defense Authorization Act for Fiscal Year 2020*, Pub L No 116-92, §§ 5709, 5724, 133 Stat 1790 (2019).

52 The Senate Select Committee on Foreign Interference through Social Media was established in 2019: 'Select Committee on Foreign Interference through Social Media', *Parliament of Australia* (Web Page)

to investigate foreign interference,[53] and passed sweeping new laws to target the same in state governments and at universities.[54] Meanwhile, the link between foreign interference and political deepfakes has been highlighted by academic commentators in submissions to both parliamentary and departmental inquiries.[55] Commentators have also highlighted the need for anticipatory reform, particularly given that elections generally cannot be 'redone' without overcoming significant legal hurdles.[56] In the absence of a new election, there is no practical remedy a court could offer post-election once a deepfake has been viewed. Reform is therefore needed *prior* to any impact on an Australian election. This is especially the case as the use of deepfakes may benefit a particular political party (whether or not they supported the use of the technology) and that party may then be unwilling to support a review into the impact of deepfake technology on their electoral victory.

### D The Need for Law to Capture (and Combat) Political Deepfakes

Protection against deepfakes cannot be left to the social media platforms on which they are shared. While some platforms have developed policies to combat deepfakes,[57] this type of remedy is insufficient for three reasons. First, even where a video is removed by the platform this does not necessarily counter the harm, and without legal powers to compel the social media platforms, an affected party cannot seek a retraction or public recognition that the video was fake. Second, not all social media companies' current disinformation policies address deepfakes, nor is there a guarantee that existing policies are sustainable. Third, definitions of 'deepfake' may vary between social media platforms and may not capture *all* videos that have been edited to mislead viewers – for example, current disinformation policies do not capture the Nancy Pelosi example discussed above.[58] In order to

---

<https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Foreign_Interference_through_Social_Media>.

53 Australian Security Intelligence Organisation Amendment Bill 2020 (Cth).

54 See, eg, Australia's Foreign Relations (State and Territory Arrangements) Bill 2020 (Cth); Australia's Foreign Relations (State and Territory Arrangements) (Consequential Amendments) Bill 2020 (Cth).

55 News and Media Centre University of Canberra and the Virtual Observatory for the Study of Online Networks Australian National University, Submission No 8 to Senate Select Committee on Foreign Interference through Social Media, Parliament of Australia, *Foreign Interference through Social Media* (2020) 3; The Allens Hub for Technology, Law and Innovation, Submission No 2 to Department of Foreign Affairs and Trade, Australian Government, *International Cyber and Critical Technology Engagement Strategy* (16 June 2020) 2.

56 In the US context the Supreme Court has blocked recounts in close presidential races: *Bush v Gore*, 531 US 98 (2000); Jack M Balkin, 'Bush v. Gore and the Boundary between Law and Politics' (2001) 110(8) *Yale Law Journal* 1407; Richard Posner, 'Bush v Gore: Prolegomenon to an Assessment' (2001) 68(3) *University of Chicago Law Review* 719, 736. Subsequent analysis revealed that Gore should have won Florida and the presidential election had a *state-wide* review of all contested ballots been conducted. However, this was not the remedy Gore had sought: Wade Payson-Denney, 'So, Who Really Won? What the Bush v. Gore Studies Showed', *CNN* (online, 31 October 2015) <https://edition.cnn.com/2015/10/31/politics/bush-gore-2000-election-results-studies/index.html>.

57 Aaron Holmes, 'Facebook Just Banned Deepfakes, but the Policy Has Loopholes – And a Widely Circulated Deepfake of Mark Zuckerberg Is Allowed to Stay Up', *Business Insider* (online, 8 January 2020) <https://www.businessinsider.com/facebook-just-banned-deepfakes-but-the-policy-has-loopholes-2020-1?r=AU&IR=T>.

58 'Facebook Refuses to Remove Doctored Nancy Pelosi Video', *The Guardian* (online, 4 August 2020) <https://www.theguardian.com/us-news/2020/aug/03/facebook-fake-nancy-pelosi-video-false-label>.

ensure consistent, and therefore fair, treatment of political deepfakes, measures must be captured in law rather than left to discretionary company policy. This approach also ensures that Parliament can set appropriate limits on what type of videos are or are not captured by the law, and tailor appropriate exemptions.

## III EVALUATION OF PRIVATE PROTECTIONS

This Part analyses the scope of current Australian laws and regulations to combat deepfakes, and the *private* remedies that are available to the subjects of a deepfake. *Public* remedies will be discussed in Part IV. This Part explores two general areas of private law: copyright law and tort law. These feature in the bulk of analysis by US commentators who have considered the legal options currently afforded to individuals who are the subject of a deepfake. Such commentary is, however, often relatively brief, forming only a small part of a larger article.[59] Additionally, little analysis has, to date, been conducted in an Australian context.

Before embarking on this analysis, it is worth noting some general points. Intellectual property and tort law provide private remedies allowing victims to bring personal actions to have deepfakes taken down, and to seek damages for any loss or injury they have suffered. Electoral regulations, discussed in Part IV, instead form a hybrid private-public remedy given the work of both the AEC and political parties and candidates in enforcing electoral regulations. The relevance of this distinction will be discussed when analysing a possible remedy, but ultimately the identity of the person bringing the action, and the speed at which they can do so are critical in the context of political deepfakes. This is because, as adverted to above, damages are unlikely to be an appropriate remedy for cases involving political deepfakes. Instead, the preferred remedy is the removal of the deepfake in a timely manner, so as to avoid any adverse impact on a politician's performance in an election.[60] More simply put, it is impossible to put a price on political power.

### A Copyright Law

Copyright law has been suggested by some commentators as a potential solution to the threat posed by deepfakes.[61] In a recent high profile example, the US reality television stars 'the Kardashians' were successful in an action to remove a deepfake from YouTube using existing copyright infringement procedures.[62] The

59 See, eg, Edvinas Meskys et al, 'Regulating Deep Fakes: Legal and Ethical Considerations' (2020) 15(1) *Journal of Intellectual Property Law & Practice* 24, 29.

60 See, eg, the concern raised at the 2019 federal election about the use of signs that mimic AEC colours: Paul Karp, 'Oliver Yates May Take Liberals to Court of Disputed Returns over "Deceptive" Election Signs', *The Guardian* (online, 21 May 2019) <https://www.theguardian.com/australia-news/2019/may/21/oliver-yates-may-take-liberals-to-court-of-disputed-returns-over-deceptive-election-signs>.

61 Meskys et al (n 59) 29.

62 Mathew Katz, 'Kim Kardashian Can Get a Deepfake Taken off YouTube. It's Much Harder for You', *Digital Trends* (online, 17 June 2019) <https://www.digitaltrends.com/social-media/kim-kardashian-deepfake-removed-from-youtube/>. The original footage used in the video was featured in *Vogue*.

deepfake is however still accessible on other platforms including Instagram.[63] Given its potential, this section explores the application of Australian intellectual property law to political deepfakes by analysing copyright subsistence, before addressing infringement, exceptions and limitations of copyright law.

Deepfakes pose a number of challenges to copyright law, including the novel question about whether copyright would, or *should*, subsist in the final work. This is important, as, if copyright subsists in a deepfake, laws that purported to strip this copyright may raise issues surrounding the acquisition of property on just terms.[64] Laws that merely regulated the *use* of the videos would however not be limited.[65] Given the requirement for human authorship for copyright to subsist in a work under Australian copyright law,[66] it is likely that copyright would *not* currently subsist in deepfakes.[67] This does not, however, mean that creators will not be liable if they infringe on another's copyright.

### *1 Subsistence of Copyright*

In assessing whether copyright subsists in a work, a court needs to assess whether the work is original. This is a question of fact,[68] which requires courts to determine whether a *human* author exercised 'independent intellectual effort' in the production of the material work.[69] In *Telstra Corporation Ltd v Phone Directories Co Pty Ltd*, the Federal Court applied this test to a written work created through a largely automated process, finding that copyright did not subsist in the resulting work.[70] In discussing how the test applied to computer programs and automated processes, Perram J stated:

> So long as the person controlling the program can be seen as directing or fashioning the material form of the work there is no particular danger in viewing that person as the work's author. … [However] the performance by a computer of functions ordinarily performed by human authors will mean that copyright does not subsist in the work …[71]

---

63 Ibid.

64 *JT International SA v Commonwealth* (2012) 250 CLR 1.

65 The key issue being whether an interest, benefit or advantage of a proprietary nature is acquired by the Commonwealth or another party: ibid.

66 *Copyright Act 1968* (Cth) s 32(1); *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 458, 493–6 [95]–[106] (Gummow, Hayne and Heydon JJ) ('*IceTV*'); *Telstra Corporation Ltd v Phone Directories Co Pty Ltd* (2010) 194 FCR 142 ('*Phone Directories*'); Sam Ricketson, 'The Need for Human Authorship: Australian Developments: *Telstra Corp Ltd v Phone Directories Co Pty Ltd*' (2012) 34(1) *European Intellectual Property Review* 54; Dilan Thampapillai, 'If Value Then Right? Copyright and Works of Non-human Authorship' (2019) 30(2) *Australian Intellectual Property Journal* 1; Dilan Thampapillai, 'The Gatekeeper Doctrines: Originality and Authorship in the Age of Artificial Intelligence' (2019) 10 *WIPO-WTO Colloquium Papers* 1.

67 There are however open questions regarding whether the provisions of the *Copyright Act 1968* (Cth) should be amended to capture works created through automated processes. Similar amendments were made in the United Kingdom: *Copyright, Designs and Patents Act 1988* (UK) s 9.

68 *IceTV* (2009) 239 CLR 458, 494–5 [99] (Gummow, Hayne and Heydon JJ).

69 *Sands & McDougall Pty Ltd v Robinson* (1917) 23 CLR 49, 52 (Isaacs J).

70 *Phone Directories* (2010) 194 FCR 142.

71 Ibid 178–9 [118]. The other members of the Court made similar statements: see 171 [89]–[90] (Keane CJ), 191 [169] (Yates J).

To determine whether copyright subsists in a deepfake, a court will need to determine whether this test should extend to artistic works. It is likely that a court would find this to be the case. Relevantly, artistic works are afforded *less* protection than literary works in the *Copyright Act 1968* (Cth) ('*Copyright Act*'),[72] and the Act does not distinguish between literary and artistic works in terms of the requirement for originality.[73] In applying the test, a court would need to determine the extent to which a person operating a neural network to create a deepfake 'direct[ed] or fashion[ed] the material [final] form of the work'.[74] This question is complex as an individual is involved at various stages of the process, including: selecting the images used to train the neural net, deciding when the neural net is ready, and selecting the video to 'swap' the face onto. Despite this involvement, it is the trained neural net that performs most of the decision-making. It is therefore likely that in Australia, copyright *would not* subsist in a deepfake.

### *2 Copyright Infringement*

The question of copyright infringement is distinct from whether copyright subsists in a work. The *Copyright Act* prevents individuals who do not own the copyright in a particular work from 'the doing in Australia of, any act comprised in the copyright'.[75] The burden of proving infringement lies on the copyright holder. In the context of *political* deepfakes, it is likely that, in creating a deepfake, an individual will draw on news content, as this is where video and audio of politicians are most accessible. Notably, the protections afforded to television and sound broadcasts by the *Copyright Act* are not as extensive as those afforded to artistic works.[76] Nonetheless the Act still prohibits the communication of sound recordings and television broadcasts to the public.[77] This could prima facie be established where news footage was used to create a deepfake. To establish infringement, a copyright holder must prove that the works are objectively similar, there was a causal connection between the original work and the infringing work, and that a substantial part of the copyright work was infringed.[78]

How these tests will apply to deepfakes has not yet been resolved by a court or explained in existing academic literature. What is clear is that there will be significant challenges in applying the tests due to the 'black box' nature of machine-learning systems. This nature means that while the inputs to the system are known (ie, the training data and the video into which the face will be swapped), the precise steps it takes to create the deepfake are not.[79] It will therefore be unclear precisely

72 For comparison, see ss 31(1)(a), 31(1)(b).

73 Ibid s 32(1): 'copyright subsists in an original literary, dramatic, musical or artistic work'.

74 *Phone Directories* (2010) 194 FCR 142, 178 [118] (Perram J).

75 *Copyright Act 1968* (Cth) s 36(1).

76 Ibid s 87.

77 Ibid ss 85, 87.

78 See, eg, *Elwood Clothing Pty Ltd v Cotton On Clothing Pty Ltd* (2008) 172 FCR 580, 588 [41] (the Court). In assessing whether a substantial part of the work was infringed, what is relevant is the quality of the work. This requires an assessment of the independent intellectual effort put into the relevant material: *IceTV* (2009) 239 CLR 458, 479 [49]–[50] (French CJ, Crennan and Kiefel JJ).

79 This was discussed in relation to automated decision-making and the resulting transparency and accountability issues that arise: Bateman (n 1).

what material was used by the machine-learning system, or the extent to which it is replicated in the final form of the deepfake. This will pose significant challenges to copyright holders (the news companies) – especially given that they will need to demonstrate that a substantial part of *their* work was infringed. Where a deepfake swaps a face into a video clip owned by a single copyright holder this issue would not arise.[80] However, if the deepfake creator stages their own scene and merely swaps an individual's face or voice into this video (using a compilation of other copyright holders' work to perform the face swap), then establishing infringement will be complex. Further, the potential compensation that would be awarded to an individual copyright holder would likely be small,[81] making bringing an action (and bearing the resulting risk of an adverse costs order) unattractive.

### *3 Copyright Exemptions*

In addition to the challenge of establishing infringement, in certain cases, deepfake creators or distributers may be able to avail themselves of exemptions in the *Copyright Act*. The Act provides an exemption where a work is a 'fair dealing … for the purpose of parody or satire'.[82]

While courts have historically used dictionaries to aid in statutory interpretation,[83] academic commentators have suggested that a broader definition of comedy and satire would give effect to the legislative intent behind the provisions.[84] These academic commentators have suggested that 'ordinary definitions', that is, the use of comedy and satire in practice, would better achieve the stated purpose of the exemptions: promoting 'free speech and Australia's fine tradition of satire by allowing our comedians and cartoonists to use copyright material for the purposes of parody or satire'.[85] Other academics have suggested that the exemption should be read broadly, with the primary test to be applied being whether the work 'adds significant new expression so as not to be substitutable for the original work'.[86] Regardless of the approach adopted by the courts, it is likely that at least some deepfakes could fall within a comedy and satire exemption, with many of them

---

80 This is often the case for example with regard to pornographic deepfakes, where an individual's face is swapped into a video owned by a single entity.

81 This is analogous to individual copyright infringement claims against individuals who pirate movies. Collectively the action is worth bringing but where courts limit the options of copyright holders, they may abandon the action: see, eg, *Dallas Buyers Club LLC v iiNet Ltd* (2015) 245 FCR 129. While Dallas Buyers Club LLC was successful in getting preliminary discovery over IP addresses, Perram J attached conditions relating to what could be communicated to the individuals identified to limit the possibility of 'speculative invoicing': at 148–9 [83]. The court later rejected the proposed letter in *Dallas Buyers Club LLC v iiNet Limited [No 3]* (2015) 327 ALR 695.

82 *Copyright Act 1968* (Cth) s 103AA.

83 The *Macquarie Dictionary* being preferred: Michael Kirby, 'Statutory Interpretation: The Meaning of Meaning' (2011) 35(1) *Melbourne University Law Review* 113, 124.

84 Conal Condren et al, 'Defining Parody and Satire: Australian Copyright Law and Its New Exception' (2008) 13(3) *Media and Arts Law Review* 273 ('Defining Parody and Satire Part 1'); Conal Condren et al, 'Defining Parody and Satire: Australian Copyright Law and Its New Exception: Part 2: Advancing Ordinary Definitions' (2008) 13(4) *Media and Arts Law Review* 401.

85 Commonwealth, *Parliamentary Debates*, House of Representatives, 19 October 2006, 2 (Philip Ruddock, Attorney-General) quoted in Condren et al, 'Defining Parody and Satire Part 1' (n 84) 274.

86 Nicolas Suzor, 'Where the Bloody Hell Does Parody Fit in Australian Copyright Law?' (2008) 13(2) *Media and Arts Law Review* 218, 220.

made for the purpose of ridiculing or critiquing politicians using very little copyrighted material. If the broad approach is taken, deepfakes would not be viewed as 'substitutable' to the original work, with the creator/author effectively using collated images for tell their own story.

Ultimately, it is unlikely that the exemption would be determinative in the overall protection afforded by copyright law, but it is worth acknowledging that its utility would, at least in some cases, be limited by the fair dealing for comedy or satire exemption.

#### *4 Limitations of Copyright Law*

Whether an individual could prove that a deepfake infringed their copyright is uncertain given the black box nature of neural networks, and the possible application of the fair dealing exemptions. There are, however, additional limitations to the protection afforded by copyright law to political deepfakes as, quite often, the politician or political party will not be the relevant copyright holder. For example, politicians often give public speeches that are recorded by broadcasters and published online. The use of this footage to train a neural net, even if it did infringe copyright, would not provide a remedy to the politician or political party. At best, the politician could request that the relevant copyright holder(s) pursue the creator of the deepfake.

It is unclear whether media companies would be willing to pursue such action, as they suffer no real harm from the infringement, and may in fact see a benefit in terms of viewer engagement. Even if they did so, the length of this process would eliminate any utility to the politician. This is especially the case where deepfakes are published on the eve of an election. In that scenario, a politician's ability to respond to a deepfake may in fact be limited by electoral blackout laws. These laws bar television and radio electoral advertising close to elections.[87] As such, deepfakes communicated over social media would not be captured by the restrictions while politicians would be limited in how they could respond to disinformation in the deepfake. The blackout laws have previously been critiqued due to the inconsistent treatment of different forms of advertising, but amendments have not yet been proposed.[88] While, in the author's view, amendments equalising the treatment of different forms of political advertising are desirable, they will not, of themselves, address the challenge posed by political deepfakes. Further analysis of the blackout laws therefore is outside the scope of this article.

### B Tort Law

There are two potential torts that may provide a remedy to the subjects of a political deepfake: defamation[89] and passing off.[90]

87 *Broadcasting Services Act 1992* sch 2 s 3A.

88 Jordan Guiao, 'Distorting the Public Square: Political Campaigning on Social Media Requires Greater Regulation' (Discussion Paper, Australia Institute, November 2019) 5.

89 See, eg, Meskys et al (n 59) 26.

90 Emma Perot and Frederick Mostert, 'Fake It Till You Make It: An Examination of the US and English Approaches to Persona Protection as Applied to Deepfakes on Social Media' (2020) 15(1) *Journal of Intellectual Property Law & Practice* 32, 35–6.

### *1 Defamation Law*

Australian defamation law has evolved from statute passed by the New South Wales Legislative Council in 1847,[91] through to the adoption of a national uniform law.[92] This evolution has been accompanied by a significant increase in the number of defamation proceedings launched. Indeed, despite the common stereotype of the Australian larrikin, Australia is seen as the defamation capital of the world.[93] This growth has coincided with the rise of social media, and is driven by a significant number of low-value claims.[94] Given this, in terms of legal actions politicians may seek to rely on to combat deepfakes, defamation is a likely candidate. Australian politicians have regularly used defamation to try to remove content harmful to their reputations. For example, Pauline Hanson was successful in obtaining an injunction against the Australian Broadcasting Corporation preventing them from playing the satirical song 'Backdoor Man'. The injunction was upheld unanimously on appeal.[95]

Broadly, to succeed in an action for defamation, a plaintiff must prove that:

1. The material was published by the defendant;
2. It identified the plaintiff; and
3. The material is defamatory (that is, it contains one or more defamatory imputations).[96]

In relation to deepfakes, the first element will be heavily fact dependent. Where a deepfake is created and published by someone in Australia, the element will be clearly established. This may not be the case where the deepfake is created by an overseas actor. In such cases, it may be possible for an individual to bring an action against the social media platform on which the deepfake was published. Australia-based media companies have been found liable in defamation for material published to their public Facebook pages.[97] Similarly, Google has been held to be liable for defamatory material published as part of its search results.[98] This suggests that where political deepfakes defame politicians, there may already be a number of prospective defendants, including web platforms and media platforms that promulgate the content.

---

91 Paul Mitchell, 'The Foundations of Australian Defamation Law' (2006) 28(3) *Sydney Law Review* 477.

92 For discussion see Andrew T Kenyon, 'Six Years of Australian Uniform Defamation Law: Damages, Opinion and Defence Meanings' (2012) 35(1) *University of New South Wales Law Journal* 31.

93 Matt Collins, 'Nothing to Write Home about: Australia the Defamation Capital of the World' (Speech, National Press Club, 4 September 2019). For analysis of the growth of low-scale cases see, eg, Centre for Media Transition, 'Trends in Digital Defamation: Defendants, Plaintiffs, Platforms' (Report, University of Technology Sydney, 2018) <http://s3.amazonaws.com/arena-attachments/1918329/e636f1839b7687241f593933d2770018.pdf?1521525181>.

94 Centre for Media Transition (n 93). Recent amendments to defamation laws passed in some states aim to reverse this trend; however their impact is yet to be seen: see, eg, Defamation Amendment Bill 2020 (NSW). For discussion about the laws, see Michaela Whitbourn, 'Uniformity at Risk as Defamation Reforms Set to Start in Three States on July 1', *Sydney Morning Herald* (online, 1 April 2021) <https://www.smh.com.au/national/uniformity-at-risk-as-defamation-reforms-set-to-start-in-three-states-on-july-1-20210401-p57fu5.html>.

95 *Australian Broadcasting Corporation v Hanson* [1998] QCA 306.

96 *Radio 2UE Sydney Pty Ltd v Chesterton* (2009) 238 CLR 460, 467 (French CJ, Gummow, Kiefel and Bell JJ).

97 *Fairfax Media Publications Pty Ltd v Voller* [2020] NSWCA 102. The NSW Court of Appeal decision was upheld on appeal by the High Court: *Fairfax Media Publications Pty Ltd v Voller* [2021] HCA 27.

98 *Defteros v Google LLC* [2020] VSC 219 ('*Defteros*').

Critically, as intent is irrelevant, defamation can be established even where '[t]he communication … [is] unintentional, and the publisher … [is] unaware of the defamatory matter'.[99] While the defence of innocent dissemination may apply, such a defence was found not to be available with respect to material published by Google in their image and text search results after it was made aware that such material was produced by its search results.[100]

How a court would apply these principles to a question concerning a political deepfake is uncertain, especially in circumstances where a media platform was unaware the video was fake (and therefore defamatory). Such a question will be significantly affected by proposed (but not yet introduced reforms) to defamation law to limit the liability of media companies for defamation.[101] If such laws are passed, then individuals or political parties impacted by deepfakes created by overseas actors may lack any remedy under defamation law.

The second and third elements would be easy to establish in relation to political deepfakes. This is because an ordinary reasonable person would likely believe a deepfake video portrayed the individual depicted, even where slight imperfections were present. This accords with previous judicial reasoning concerning doctored images, which were of a significantly lower quality than is achievable in a deepfake.[102] Finally, given that the purpose of using a political deepfake is to lower the likelihood of an individual voting for a particular individual or party it is probable that in many cases a deepfake would contain a defamatory imputation. However, where a deepfake was *only* targeted at a political party it would fall outside the protection afforded by defamation law – which only protects the reputation of natural persons.

### *2 Passing Off*

The classical elements of the tort of passing off under Australian law are drawn from the United Kingdom ('UK') case of *Reckitt & Colman Products Ltd v Borden Inc* ('*Reckitt & Colman*').[103] The broad test requires the establishment of the 'classical trinity', the elements of which are:

1. Reputation within Australia;
2. Misrepresentation; and
3. Damage.[104]

---

99 *Lee v Wilson* (1934) 51 CLR 276, 288 (Dixon J).

100 *Defteros* [2020] VSC 219, [134] (Richards J).

101 Michael Douglas, 'Australia's Proposed Defamation Law Overhaul Will Expand Media Freedom – But at What Cost?', *The Conversation* (online, 1 December 2019) <https://theconversation.com/australias-proposed-defamation-law-overhaul-will-expand-media-freedom-but-at-what-cost-128064>. Reforms to limit liability of media companies and intermediary platforms are currently being considered by government: see Attorneys-General, 'Review of Model Defamation Provisions: Stage 2' (Discussion Paper, 2021) <https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/discussion-paper-stage-2.pdf>.

102 See, eg, *Hanson-Young v Bauer Media Ltd [No 2]* [2013] NSWSC 2029.

103 *Reckitt & Colman Products Ltd v Borden Inc* [1990] 1 WLR 491 ('*Reckitt*'). *Reckitt* was applied by the High Court in *ConAgra Inc v McCain Foods (Aust) Pty Ltd* (1992) 33 FCR 302 ('*ConAgra*').

104 *Reckitt* [1990] 1 WLR 491, 499 (Lord Oliver); *ConAgra* (1992) 33 FCR 302, 355–6 (Gummow J).

In Australia, the *Reckitt & Colman* test has been regularly used to protect celebrities' images where individuals or businesses have implied that their goods or services have been approved or endorsed by the celebrity. For example, Ita Buttrose was successful in recovering damages where her image was used in a false endorsement.[105] Similarly, Paul Hogan was successful in recovering damages where an advertisement used an actor dressed in similar attire to his costume in Crocodile Dundee and used the now-famous line 'that's not a knife'.[106] This suggests that (similar to the analysis above in terms of defamation law) a deepfake could meet the requirements of this test, even if it contains slight glitches or imperfections. This is because courts are not assessing whether an individual is likely to believe that the celebrity portrayed really did say the words attributed to them, but instead whether an individual would form a connection in their mind such that they would believe 'the goods are … endorsed by the [celebrity]'.[107] In contrast, UK courts have historically been less willing to extend the doctrine of passing off beyond its traditional business roots,[108] although this has recently begun to shift.[109]

In the case of political deepfakes, the critical issues are whether a subject had a significant enough reputation in Australia, and whether a misrepresentation in the *commercial* sense protected by the tort had occurred. This case would differ from the traditional endorsement cases discussed above, as it is unlikely that a political deepfake would be used to advance a business interest. Instead, the deepfake would likely target a political interest: to affect public opinion regarding a politician, or the platform of a given politician or party. This analysis is analogous to the position adopted by Perot and Mostert who suggested that passing off may afford protections to individuals for certain categories of deepfakes in the UK.[110] The authors did not discuss the application of the test to political deepfakes. Where an opposing political party utilises a deepfake to further their political interests, this link may be easier to establish. In most cases involving political deepfakes, however, the current test for passing off is unlikely to serve as a suitable protection.

### *3 Limitations of Tort Law*

As outlined in the above analysis, the efficacy of either defamation or passing off in combatting political deepfakes is limited. In addition to the gaps identified above, the primary limitation of tort law pertains to the remedies available to an aggrieved plaintiff. While courts are able to grant injunctions to prevent ongoing

105 *Buttrose v The Senior's Choice (Australia) Pty Ltd* [2013] FCCA 2050 ('*Buttrose*').

106 *Pacific Dunlop Ltd v Hogan* (1989) 23 FCR 553. Hogan has been an active celebrity in this space, also bringing an action against a company selling a 'Crocodile Dundee Koala Bear': *Hogan v Koala Dundee Pty Ltd* (1988) 83 ALR 187. See also 'Grill'd Settles Dispute with Paul Hogan', *SBS News* (online, 5 February 2018) <https://www.sbs.com.au/news/grill-d-settles-dispute-with-paul-hogan>.

107 *Buttrose* [2013] FCCA 2050, [48] (Jones J).

108 See, eg, *Elvis Presley Trade Marks* [1999] RPC 567, 598 (Brown LJ): 'there should be no … assumption that only a celebrity … may ever market … [their] own character'.

109 See *Irvine v Talksport Ltd [Nos 1 and 2]* [2003] 2 All ER 881; *Fenty v Arcadia Group Brands Ltd* [2015] EWCA Civ 3.

110 Perot and Mostert (n 90) 35–6.

damage, their use is limited.[111] This is especially the case for interlocutory applications where a court will only interfere in exceptional cases.[112] The reasons for this were summarised by the Federal Court in *Rush v Nationwide News Pty Ltd [No 9]*:[113]

> There are essentially three reasons why caution is warranted … [first that] free speech might be unnecessarily curtailed or restricted … [second that] it is not known whether publication of the matter will in fact invade the legal right of the applicant; and third, the fact that the defence of justification is ordinarily a matter for decision by a jury, not by a judge sitting alone …[114]

Additionally, defamation cases – the more useful remedy for individual politicians – are extremely costly and lengthy to run. Indeed, costs have been estimated to be as high as $80,000–$100,000 for cases involving only $10,000 in damages, leading to the introduction of legislation that would have removed the ability of parties to recover costs in low-value matters.[115] The cost-benefit analysis in the case of a deepfake affecting only 100–200 votes may be against bringing an action. Similarly, as a deepfake can be generated in a matter of days, a politician who embarked on a 'defend all cases' strategy may find themselves endlessly appearing in court. Fatigue, or mounting costs, would likely force the end to such action. In essence, the actions are limited by their personal nature, and the fact that parties may struggle to seek an injunction to prevent the ongoing harm.

Further, an award of damages would do little to restore trust in political and democratic institutions. Indeed, bringing an action can lead to increased media focus on the defamation case itself, allowing the allegedly defamatory claims to spread further. A more appropriate solution may be to empower impartial actors to secure the integrity of the voting process.

### C Summary of Applicable Private Law

As outlined above, the remedies available in private law with respect to political deepfakes are insufficient. In particular, copyright law will only protect the relevant copyright holders – who are more likely to be media companies than the politicians impacted. Additionally, even where media companies were inclined to bring an action, the black box nature of deepfake technology would make identifying whose copyright had been infringed impossible in many cases. While defamation law would provide politicians with the strongest remedy, the time and costs needed to bring a defamation action limit its utility. Similar issues pervade the tort of passing off. Ultimately, rather than a private law action for damages,

111 See, eg, *Australian Broadcasting Corporation v O'Neill* (2006) 227 CLR 57, 66 [16] (Gleeson CJ and Crennan J).

112 Benedict Bartl and Dianne Nicol, 'The Grant of Interlocutory Injunctions in Defamation Cases in Australia following the Decision in *Australian Broadcasting Corporation v O'Neill*' (2006) 25(2) *University of Tasmania Law Review* 156.

113 [2019] FCA 1383.

114 Ibid [8] (Wigney J).

115 New South Wales, *Parliamentary Debates*, Legislative Assembly, 18 September 2003, 3586–7 (David Barr). The laws were not passed in 2003; however, a Bill that will likely have a similar effect has now been passed in some states: see, eg, Defamation Amendment Bill 2020 (NSW).

those impacted by a deepfake likely want a 'public law' protection allowing them to take down harmful deepfakes.

## IV EVALUATION OF PUBLIC LAW PROTECTIONS

Federal elections are governed by the *Commonwealth Electoral Act 1918* (Cth) ('*Electoral Act*'). While some Australian states have moved to prohibit specific uses of deepfake technology, notably in the context of intimate partner violence,[116] there are no specific laws or regulations concerning their use in federal, state or local elections.[117] Instead, the *Electoral Act* creates a number of general electoral offences that *may* apply to political deepfakes.[118] Where an offence has occurred, the *Electoral Act* creates a hybrid public-private enforcement regime, with both the AEC and candidates in an election able to seek an injunction to prevent conduct that would contravene the *Electoral Act*.[119] While there is some controversy concerning the availability of general administrative review rights against the AEC,[120] this question is not concerned with jurisdiction over electoral offences.[121] Therefore, while it remains unclear what remedies, if any, a private citizen has under the *Electoral Act*, this question is beyond the scope of this article although exploration of that topic may yield additional (and novel) remedies to the challenges posed by political deepfakes.

Relevantly, if requested by a candidate (during an election period), or the AEC, the Federal Court may grant an injunction where an offence has occurred or is likely to occur 'if in the opinion of the [Court] it is desirable to do so'.[122] Therefore, if the publication or distribution of a deepfake contravened a section of the *Electoral Act*, a court would be able to prohibit its publication through an injunction. This is exactly the remedy that the subject of a deepfake would be likely to seek. The below analysis highlights how two relevant offences would apply to political deepfakes.

### A Misleading and Deceptive Conduct

Section 329 of the *Electoral Act* creates an offence for misleading and deceptive publication, which, on its face, would appear to apply to political deepfakes. The offence is however limited in its application. Section 329 relevantly states:

---

116 *Crimes Act 1900* (NSW) ss 91N, 91Q. The use of deepfake technology would fall within the definition of 'altered image'.

117 The latter two are beyond the scope of this article; however, state electoral regulations would provide guidance if they regulated deepfakes.

118 *Electoral Act 1918* (Cth) pt XXI.

119 Ibid s 383.

120 Graeme Orr, 'Judicial Review of Electoral Affairs' (Conference Paper, AIAL National Administrative Law Forum, July 2011). See also Graeme Orr and George Williams, 'Electoral Challenges: Judicial Review of Parliamentary Elections in Australia' (2001) 23(1) *Sydney Law Review* 53.

121 Orr (n 120).

122 *Electoral Act 1918* (Cth) s 383(1).

> **329 Misleading or deceptive publications etc.**
> (1) A person shall not, *during the relevant period* in relation to an election under this Act, print, publish or distribute, or cause, permit or authorize 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*.[123]

While 'matter or thing' would likely include deepfake videos, and the term 'publish' includes distribution over the internet,[124] section 329 would be of limited use for two reasons. First, the section only applies during the relevant period – which is defined under the *Electoral Act* to be the period from the issue of writs to the conclusion of the election.[125] This means that the section would not apply to any communications or materials before the issuing of the writs. This limitation is not, however, critical. As noted above, the primary concern regarding deepfakes is their release close to an election where insufficient time remains to verify whether the contents of the video are true. As such, the limitation of section 329 to the time between the issue of writs and the end of the election would not be fatal to its use. More significant, however, is the limitation of the section to conduct 'in relation to the casting of a vote'. Courts have consistently held that this language limits section 329 to only apply to cases where the misleading or deceptive conduct relates to *how* an elector (having already decided who will be receiving their vote) would number the boxes on a ballot paper.[126] For example, the Full Federal Court in *Garbett v Liu*[127] stated:

> The provision is not concerned with a matter or thing which is misleading or deceptive and which might influence an elector in forming a judgment … It is concerned with the casting of the vote … The distinction is one between the formation of the political or voting judgment of the elector, and *its recording or expression*.[128]

Section 329 therefore does not guard against misleading or deceptive conduct in relation to electoral choices.[129] This can be contrasted with various state and territory electoral Acts which contain (or will soon contain)[130] prohibitions on false and misleading statements in advertising. For example, the South Australian *Electoral Act* creates an offence where:

> A person who authorises, causes or permits the publication of an electoral advertisement (an advertiser) is guilty of an offence if the advertisement contains a statement purporting to be a statement of fact that is inaccurate and misleading to a material extent.[131]

---

123 Ibid s 329(1) (emphasis added).
124 Ibid s 329(6).
125 Ibid s 322.
126 See, eg, *Evans v Crichton-Browne* (1981) 147 CLR 169.
127 (2019) 273 FCR 1.
128 Ibid 8 [31], 10 [36] (emphasis added).
129 Historically, the section *did* engage with generally misleading and deceptive conduct – but the former provision was repealed: George Williams, 'Truth in Political Advertising Legislation in Australia' (Research Paper No 13, Parliamentary Library, Parliament of Australia, 24 March 1997).
130 *Electoral Amendment Act 2020* (ACT) s 13, which will insert a new section 297A into the *Electoral Act 1992* (ACT).
131 *Electoral Act 1985* (SA) s 113(2).

This provision, as of 2019, was the strongest 'truth in political advertising' law globally.[132] Notably, the University College London Report, in making this finding, outlined that amendments to the South Australian legislation in 1997 allowing the Electoral Commissioner to intervene to request an advertisement be immediately withdrawn meant that action could be taken before 'the election was over'.[133]

The utility of the South Australian provision has, however, been called into question.[134] For example, a former South Australian Electoral Commissioner outlined to a Federal parliamentary inquiry that:

> [H]e did not believe the South Australian legislation had had any appreciable effect on the nature of electoral advertising in the State. Instead, he considered that the legislation opened up opportunities for individual candidates to disrupt the electoral process by lodging nuisance complaints.[135]

Additionally, as the South Australian and Australian Capital Territory provisions apply only to *paid* advertising, they would not cover the use of deepfakes spread through social media by individuals not connected to a political campaign.

Nevertheless, absent such a provision, at a federal level, a deepfake falsely showing a candidate engaging in criminal activity, or outlining a false policy position which may mislead a voter as for whom they *wish* to vote would not be captured through the operation of section 329. In contrast, section 329 would prohibit the creation of a deepfake which, for example, falsely suggested which box a voter should number if they wished to vote for a particular party.[136]

### B Publication of Matter regarding Candidates

The second provision that, on its face, appears to apply to political deepfakes is section 351, which relevantly states:

> **351 Publication of matter regarding candidates**
>
> (1) If, in any matter announced or published by any person, or caused by any person to be announced or published, on behalf of any association, league, organization or other body of persons, it is:
>
> (a) claimed or suggested that a candidate in an election is associated with, … that association, league, organization or other body of persons; or
>
> (b) expressly or impliedly advocated or suggested:
>
> (i) … that a voter should place in the square opposite the name of a candidate on a ballot paper a number not greater than the number of Senators to be elected; or

132 Alan Renwick and Michela Palese, 'Doing Democracy Better: How Can Information and Discourse in Election and Referendum Campaigns in the UK Be Improved?' (Report, University College London, March 2019) 22.

133 As was the case where courts had to make a determination: ibid 23.

134 Ibid.

135 Senate Standing Committee on Finance and Public Administration, Parliament of Australia, *Inquiry into Bills Concerning Political Honesty and Advertising* (Report, August 2002) 88 [5.60].

136 This is analogous to the creation of a false how-to-vote card, which the AEC has stated would be captured by the section: Australian Electoral Commission, Submission No 1 to Joint Standing Committee on Electoral Matters, Parliament of Australia, *Inquiry into Allegations of Irregularities in the Recent South Australian State Election* (June 2010) 2–3.

> (ii) … that that candidate is the candidate for whom the first preference vote should be given;
>
> that person commits an offence.

A survey of results from two databases was not able to find any cases where the section has been used.[137] However, the section does appear to prohibit certain types of political deepfakes. This is because a deepfake of a candidate speaking may suggest to viewers that they hold the views outlined in the video. A key limitation of the provision is that the deepfake would have to be published *on behalf of* an organisation (or the associated terms used in the *Electoral Act*). The deepfake would then also have to suggest that the candidate is linked to the organisation, or suggest to voters how they should number their ballot paper (this part of section 351(1)(b) is similar to section 329). While it would be possible for a deepfake to fall within the section, it would be straightforward to design a deepfake to avoid such an outcome. Similarly, the section would not prohibit an individual, of their own volition, creating or disseminating political deepfakes.

### C Summary of Applicable Public Law

As the above section has outlined, there are only limited public law protections available to political actors or the AEC as a means of pursuing those responsible for political deepfakes. While some limited types of deepfake will be captured, sophisticated actors will be able to avoid the subject matter areas that may run afoul of electoral regulation. The lack of remedy creates a gap in the law highlighting that the current legal framework is not fit for purpose at least insofar as it deals with the threat posed by political deepfakes.

## V PROPOSED REFORM

Given the analysis above, reform is needed to combat political deepfakes. The following section discusses the constitutional limitations that would apply to federal laws developed to combat the threat of political deepfakes, before outlining a proposed model law.

### A Commonwealth Powers

The Federal Government has a wide array of constitutional heads of power to draw on to regulate against the creation or distribution of *political* deepfakes. For example, the Commonwealth has the power to legislate with respect to elections,[138]

137 With the usual caveats around use of available databases, the search terms "Electoral Act 1918 (Cth)" AND "351" AND "misleading" were used across two databases. No relevant cases were found.

138 *Constitution* s 51(xxxvi).

copyright,[139] telecommunications,[140] corporations,[141] defence[142] and external affairs.[143] In combination these powers would likely allow[144] the Commonwealth to:

1. regulate the creation and content of political deepfakes by political parties or related entities within the context of federal elections using the elections power;
2. extend current copyright law to prohibit the creation of deepfakes;
3. ban the distribution of political deepfakes within and outside an electoral period through the use of a carriage service (including the internet);[145]
4. create offences relating to the creation or dissemination of deepfakes for the purpose of influencing elections due to the threat they pose to security;
5. impose duties on corporations acting in Australia to prevent the distribution of political deepfakes;[146] and
6. extend any offence provisions overseas.[147]

Given the wide array of options identified above, the key question to answer in determining what can be done to regulate against the threats identified in Part II is what limits, if any, the *Constitution* imposes with respect to these laws. Given the focus of this article on *political* deepfakes, the relevant limit is the operation of the implied freedom of political communication ('IFPC').

## B Limits Imposed by the IFPC

The IFPC is a limitation on legislative and executive power derived from the text and structure of the *Constitution*.[148] The current test was applied by a majority

---

139 Ibid s 51(xviii).

140 Ibid s 51(v).

141 Ibid s 51(xx).

142 Ibid s 51(vi).

143 Ibid s 51(xxix).

144 The list is not intended to be an exhaustive statement regarding government power, merely to provide several examples identified by the author. No comment is made regarding the desirability of these regulations.

145 This could likely be done through the telecommunications powers under which similar regulations barring the dissemination of child exploitation material have been passed: see *Criminal Code Act 1995* (Cth) sch div 474 sub-div D ('*Commonwealth Criminal Code*').

146 This could be done using the corporations power contained in section 51(xx) of the *Constitution*, and would mirror current laws regarding child exploitation material: for discussion, see below n 171 and accompanying text.

147 This could be done using the external affairs power in section 51(xxix) of the *Constitution*, analogous to current foreign interference laws: *Foreign Influence Transparency Scheme Act 2018* (Cth) s 7. Albeit the utility of such laws would be questionable, as foreign states can limit the utility of prosecution by not allowing their citizens to be extradited: see, eg, Amy Maguire, 'MH17 Charges: Who the Suspects Are, What They're Charged With, and What Happens Next', *The Conversation* (20 June 2019) <https://theconversation.com/mh17-charges-who-the-suspects-are-what-theyre-charged-with-and-what-happens-next-119155>. Notably both Russia and the People's Republic of China (nations which have been condemned internationally for their foreign interference efforts) have domestic laws that would prevent Australia from seeking extradition of their nationals: article 61 of the *Constitution of the Russian Federation*; «中华人民共和国引渡法» [Extradition Law of the People's Republic of China] (People's Republic of China) National People's Congress, Order No 42, 28 December 2000, art 8.

148 *Australian Capital Television Pty Ltd v Commonwealth* (1992) 177 CLR 106. For discussion, see *Comcare v Banerji* (2019) 267 CLR 373, 395 [20] (Kiefel CJ, Bell, Keane and Nettle JJ) ('*Banerji*').

of the High Court in *McCloy v NSW*[149] and further clarified in *Brown v Tasmania*.[150] It requires a court to answer three questions:

1. Does the law effectively burden the implied freedom … ?
2. … is the purpose of the law legitimate … ?
3. … is the law reasonably appropriate and adapted to advance that legitimate objective … ?[151]

In assessing this third question a court must consider whether the law is suitable, necessary and adequate in its balance.[152] If question (1) is answered in the affirmative and either of questions (2) or (3) are answered in the negative the law will be invalid.[153]

In terms of regulating political deepfakes, the first question a court would need to assess is whether deepfakes are political speech. If not, then the IFPC would not apply. The key issue here is whether the IFPC protects *false* speech. While it does not appear that the High Court has made a direct finding on this issue, comments in obiter from both the High Court and the South Australian Supreme Court support the proposition that false speech is protected. This is especially the case where the speech is related to a core political matter. For example, in *Roberts v Bass*[154] Gaudron, McHugh and Gummow JJ stated that defamation (which inherently is concerned with untrue statements) is limited by the IFPC.[155] A similar finding was made by the Court in *Lange v Australian Broadcasting Corporation*.[156] Even in cases concerned with false statements, courts have stepped through the entirety of the *McCloy* test to assess whether a law is adequate in its balance.[157] For example, in *Cameron v Becker*, in holding that section 113 of the South Australian *Electoral Act* did not breach the IFPC, Olsson J (with whom Bollen J agreed) appeared to hint that false speech would not attract the protection of the IFPC.[158] However, Olsson J went on to assess whether the law was 'reasonably appropriate and adapted'.[159]

This approach prevents courts from unnecessarily assessing whether speech is or is not true. Especially within the context of elections, Australian courts have taken care when applying the IFPC. For example, Kirby J in *Roberts v Bass* stated:

> Because this is the real world in which elections are fought in Australia, any applicable legal rule … must be fashioned … to reflect such electoral realities. Otherwise, before or after the conduct of elections, attempts will be made to bring to courts of law, under the guise of legal claims, the very disputes that it was the

149 *McCloy v NSW* (2015) 257 CLR 178 ('*McCloy*').
150 *Brown v Tasmania* (2017) 261 CLR 328.
151 *Clubb v Edwards* (2019) CLR 171, 186 [5] (Kiefel CJ, Bell and Keane JJ). See also *Banerji* (2019) 267 CLR 373, 398–400 [29]–[32] (Kiefel CJ, Bell, Keane and Nettle JJ).
152 *Brown v Tasmania* (2017) 261 CLR 328, 368 (Kiefel CJ, Bell and Keane JJ), 376 (Gageler J), 416–17 (Nettle J), 476–7 (Gordon J); *Banerji* (2019) CLR 373, 400 [32] (Kiefel CJ, Bell, Keane and Nettle JJ).
153 *McCloy* (2015) 257 CLR 178, 193–5 [2]–[3] (French CJ, Kiefel, Bell and Keane JJ).
154 (2002) 212 CLR 1.
155 Ibid 40–1 [102].
156 (1997) 189 CLR 520.
157 In the context of the earlier tests predating the *McCloy* test, see, eg, *Cameron v Becker* (1995) 64 SASR 238, 248 (Olsson J, Bollen J agreeing at 239).
158 Ibid 247.
159 Ibid 248.

> purpose of the representative democracy, established by the *Constitution*, to commit to the decision of the electors.[160]

This approach balances the need for laws to comply with the IFPC with the risk that courts could become an electoral and political battleground, subverting the will of the people. The United States, in contrast, has a very litigious electoral system with state and federal courts often called on to settle political controversies around voting rights, access to voting and the legitimacy of electoral results. This approach culminated in the *Bush v Gore* decision where the Supreme Court split on party lines to elect George W Bush as President.[161]

While deepfakes are a form of *false* speech in that they portray individuals making false statements, there are also many legitimate uses of deepfakes as discussed above. Deepfakes can be used as a form of parody or satire, or to educate the general population about the threat of fake news. Additionally, deepfakes can be used by politicians to make videos of *themselves* speaking in different languages in efforts to appeal to a greater share of the voting base. Laws that purport to prohibit the creation or dissemination of political deepfakes would impact on these *legitimate* uses of the technology. As members of the High Court have recently made clear, laws which impact on future communications may have a significant chilling effect.[162]

As such, and especially given the statements in *Cameron v Becker*, it seems likely that a court would find that laws that limit the publication and dissemination of deepfakes are a burden on the implied freedom. Therefore, any laws prohibiting the creation or dissemination of a political deepfake would need to be compatible with the system of representative government, and be reasonably and appropriately adapted to that legitimate purpose. The purpose underlying the laws has been addressed earlier in this article; however, in sum, the laws would aim to safeguard Australian elections and ensure that voter preferences were not subverted by deepfakes. This purpose aims to strengthen legitimate political communication and protect elections from both foreign interference and domestic threats and would likely be compatible with Australia's system of representative government. The key issue in designing such laws is therefore in ensuring that the laws are suitable, necessary and adequate in their balance. This analysis will depend on the specific measures adopted and will accordingly be discussed further below alongside the proposed legislative scheme.

## C Possible Reforms

As outlined above, there are several potential avenues for reform. In determining which approach should be taken, three questions need to be answered:

---

160 *Roberts v Bass* (2002) 212 CLR 1, 63 [172].

161 See, eg, *Bush v Gore*, 531 US 98 (2000).

162 See, eg, *LibertyWorks Inc v Commonwealth* [2021] HCA 18, [95] (Gageler J) (noting that his Honour was in dissent on this issue with the plurality finding that there was no such acceptance of strict scrutiny for prior restraint: at [50] (Kiefel CJ, Keane and Gleeson JJ)). Regardless, however the burden is analysed it is clear that laws which impact on an individual's ability to communicate about politicians using deepfakes will likely fall afoul of the first element of the *McCloy* test.

1. On whom should obligations be imposed?
2. What type of remedy is appropriate?
3. Who should be able to seek the remedy?

In answering these questions, it is important to outline the purpose of these proposed reforms: to safeguard *elections* by preventing voters from being swayed by misinformation and disinformation. While ordinarily a certain amount of misinformation is anticipated in the context of a contested election campaign, the need to combat deepfakes has been clearly articulated. The situation can be distinguished from false claims generally as there is no practical way for the subject of a deepfake to correct the record. Either people will believe the video is real, or they will not. This is different, for example, from false advertising regarding death taxes[163] or Medicare funding,[164] as these policy-based arguments can, at least in theory, be debated and corrected on the public record.[165]

In contrast, the difficulty in disproving a deepfake and the increasing ease of deepfake creation[166] justifies intervention. Care, however, must be taken to ensure that any new regulations are not used by political parties to decide electoral contests through litigation.[167] Such an outcome could erode the trust of electors in elections, and undermine the separation of powers in Australia by giving courts the ability to decide electoral contests.[168] It would also mean the law would be more likely to breach the IFPC as not being reasonably and appropriately adapted to the purpose it is seeking to achieve. Finally, the outcome could also increase the political profile of federal courts, and increase the influence of a judge's political persuasion in appointment decisions. The dangers of this potential outcome are on full display in the US, where the confirmation of Justice Amy Coney Barrett took place eight days before the election, with President Trump admitting he hoped the appointment would establish a sympathetic bench to rule upon electoral issues such as mail voter fraud.[169]

---

163 Katharine Murphy, Christopher Knaus and Nick Evershed, '"It Felt Like a Big Tide": How the Death Tax Lie Infected Australia's Election Campaign', *The Guardian* (online, 8 June 2019) <https://www.theguardian.com/australia-news/2019/jun/08/it-felt-like-a-big-tide-how-the-death-tax-lie-infected-australias-election-campaign>.

164 See, eg, Nicholas Reece, 'Why Scare Campaigns Like "Mediscare" Work: Even if Voters Hate Them', *The Conversation* (online, 14 July 2016) <https://theconversation.com/why-scare-campaigns-like-mediscare-work-even-if-voters-hate-them-62279>.

165 For example, while Labor acknowledged the impact of the 'death tax' ads on its campaign, its report into the 2019 election admits that much of the blame lay with an unwieldy policy platform and an inability to respond to the claims in a way that voters could understand: see, eg, Emerson and Weatherill (n 38) 19, 74.

166 For discussion, see Part II.

167 *Roberts v Bass* (2002) 212 CLR 1, 63 [172] (Kirby J).

168 While currently the Court of Disputed Returns can void an election, the grounds on which they can do so are extremely limited: *Electoral Act 1918* (Cth) pt XXII. Such powers have never been used.

169 Jordyn Phelps, 'Trump Argues His Nominee Needed on Supreme Court in Time to Vote on Election Legal Challenges', *ABC News* (online, 24 September 2020) <https://abcnews.go.com/Politics/trump-argues-nominee-needed-supreme-court-time-vote/story?id=73192756>. See also ABC News, 'Donald Trump's Nominee Amy Coney Barrett Confirmed to Supreme Court of the United States', *ABC News* (online, 27 October 2020) <https://www.abc.net.au/news/2020-10-27/amy-coney-barrett-confirmation-senate-supreme-court-donald-trump/12815614>.

### *1 Where Should Obligations Fall?*

This question is likely the most contentious of the three, especially given recent attempts by the Commonwealth to impose obligations on social media companies to pay for news have led to threats by social media and internet companies to withdraw or limit their Australian operations.[170] What is clear is that individuals or political parties who share deepfakes with the intention of impacting elections should be captured by the regulations. The laws should also account for scenarios where the precise author of the deepfake remains unknown (at least when action is commenced). As highlighted above, it is possible for anonymous actors overseas to be responsible for the creation and dissemination of deepfakes. What is less clear is how to regulate news media companies and social media platforms who may unknowingly assist in the distribution of a deepfake. In the context of terror attacks or child exploitation material, obligations have been imposed (both in Australia and internationally) on media companies to prohibit the sharing or uploading of content.[171] This has led to platforms creating automated tools that flag and then delete any such content.[172] While some commentators have suggested imposing obligations on social media platforms prohibiting the spread of fake news, even in the context of deepfakes, such an approach may be unwieldy and overbroad.[173] This is because defining misinformation and disinformation is much harder than defining child exploitation material, or abhorrent violent material, and as such additional content may be captured by automated detection tools (and unnecessarily censored).[174]

Obligations imposing significant penalties on service providers or content hosts where their platform is used to access that material may therefore lead to unnecessary restrictions on free speech, with providers removing more content than necessary. For example, videos that were clearly identified as deepfakes and were uploaded for educational purposes may be removed by risk averse companies using automated tools to detect and remove *all* deepfakes. This in turn would hurt the democratic process by unnecessarily restricting political communication.

To balance the need for media freedom (and avoid unnecessarily burdening social media platforms), a two-pronged approach could be used. At first instance, action could be taken against the original creator or disseminator of the deepfake. Then, if that action is successful, obligations could be imposed on social media platforms

---

170 Matthew Doran and Jordan Hayne, 'Facebook Threatens to Ban Australians from Sharing News after Google Launches Attack on Government Plans', *ABC News* (online, 1 September 2020) <https://www.abc.net.au/news/2020-09-01/facebook-threatens-to-ban-australians-from-sharing-news-content/12616216>.

171 Further measures were introduced following the live-streaming of the Christchurch terror attack to insert (among other provisions) sections 474.33 and 474.34 into the *Commonwealth Criminal Code*: Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019 (Cth). Sections 474.33 and 474.34 drew on the approach in section 474.25 which imposes obligations where an internet service provider or content host is aware that the service can be used to access child abuse material to impose similar obligations with respect to abhorrent material.

172 Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7(1) *Big Data & Society* 1.

173 Ibid 10–12.

174 Ibid.

and news companies to prevent them from *knowingly* allowing this content to be shared. This would have the effect of reducing the burden on social media companies – who would not need to decide whether material did or did not need to be removed at the first instance. They would instead be able to rely on a court determination and *then* use automated detection tools to remove any re-uploaded deepfake videos. Such an approach mirrors that taken in relation to extremist content following the Christchurch terror attacks,[175] and circumvents much of the ongoing debate around the extent of safe harbour provisions[176] as the regulation will be limited to a defined set of videos of which technology companies are aware.

By limiting the restrictions imposed in this manner, the government could leave decisions about less harmful cases (including when content should be downgraded in searches or flagged as misleading or false) to social media companies themselves, who can manage these issues under internal policies.[177] Under this approach, the government's efforts will be tailored to focus on the greater threat posed by deepfakes, ensuring that the law is not overbroad and more likely to be held to be reasonably and appropriately adapted.

### *2 What Type of Remedy Is Needed and When Should It Be Available?*

It is clear from the preceding analysis that damages are not a sufficient remedy to combat political deepfakes. Instead, what is needed is an ongoing injunction restraining the publication or republication of the relevant political deepfake. Given that deepfakes can be easily re-uploaded, a further remedy should be available: the ability to request or compel a public correction of the record by the party responsible for publishing the deepfake. This approach mirrors that contained in the South Australian and Australian Capital Territory electoral Acts regarding false political advertising.[178] Where a public retraction is required, legislation should, as a matter of course, require the retraction be in the same form and shared as widely as the original post or video. While this power is likely already captured in the wide array of orders a court may grant under the *Electoral Act*,[179] an express statement would clearly indicate its availability and help tailor the conditions attached to the order. It is worth noting that existing provisions in the *Electoral Act* require courts

175 See, eg, *Commonwealth Criminal Code* sub-div 474(H).

176 See, eg, Peter Leonard, 'Building Safe Harbours in Choppy Waters: Towards a Sensible Approach to Liability of Internet Intermediaries in Australia' (2010) 29(3) *Communications Law Bulletin* 10; Danny Friedmann, 'Sinking the Safe Harbour with the Legal Certainty of Strict Liability in Sight' (2014) 9(2) *Journal of Intellectual Property Law & Practice* 148. This approach accords with the safe harbour scheme contained in clause 91 of schedule 5 of the *Broadcasting Services Act 1991* (Cth) which requires *knowledge* to impose liability on an internet service provider. In the author's view, safe harbour protections should generally not be afforded to internet service providers in relation to electoral offences where they are aware that the content infringes electoral law.

177 For discussion on the measures already taken by social media companies, see, eg, Emma Llansó et al, 'Artificial Intelligence, Content Moderation, and Freedom of Expression' (Working Paper, Transatlantic Working Group on Content Moderation Online and Freedom of Expression, 26 February 2020).

178 *Electoral Act 1985* (SA) s 113; *Electoral Act 1992* (ACT) s 297A.

179 *Electoral Act 1918* (Cth) s 360, noting that the section is framed as an inclusive list of powers.

to make decisions as quickly as possible given the circumstances of the case.[180] This further strengthens the appropriateness of the proposed remedy.

### *3 Who Should Be Able to Seek the Remedy?*

This third question is likely the easiest of the three to answer. In line with current practice, the hybrid public-private model created by the *Electoral Act* should be applied. This would allow candidates affected (if the video occurs during an election campaign) and the AEC to bring an action. Limiting the action to candidates only during an election period further tailors the law, as it prevents overuse of the courts for political point-scoring. Allowing the Electoral Commissioner to issue notices will enable action to be taken rapidly rather than requiring court action in every case. It will also enable the Commissioner to issue take-down notices in situations where the creator or disseminator remains anonymous and a civil action against that person may not be possible. While this alone will not resolve the issue of attribution of actions taken online, especially where actions are taken by state-sponsored actors, it will go some way to providing the Commissioner with powers to remove deepfake content. Of course, in an Australian context, current electoral laws already require the identification of the individual(s) authorising electoral communications.[181]

## D Proposed Amendments

To give effect to the above, two proposed amendments to the federal *Electoral Act* are set out below:

**Section 329A Publish or distribute altered images etc.**

(1) This section applies to altered images published by any means.
(2) A person who authorises, causes or permits the publication of any matter or thing is guilty of an offence if the matter or thing contains a statement regarding electoral matters that is inaccurate or misleading to a material extent.
(3) In prosecuting a person for an offence under this section, it is a defence if:
    (a) the person proves that they did not know and could not reasonably be expected to have known, that the matter or thing was:
        (i) likely to mislead or deceive an elector to a material extent; or
        (ii) an altered image; or
    (b) the person proves that:
        (i) the material or thing was published for the purpose of education, comedy, or satire; and
        (ii) the material or thing was identified as an altered image.

180 Ibid s 363A.

181 Ibid pt XXA. One possible alternate to the proposal in Part V(D) would be to require an authorised individual to be nominated for every political deepfake published in Australia and to take down any deepfakes that are not authorised; however, such a measure would have far greater impact on political communication and accordingly in the author's view this approach is likely more proportionate to the threat it seeks to prevent.

(4) If the Electoral Commissioner is satisfied, on the balance of probabilities, that a matter or thing has been published in relation to electoral matters that is inaccurate or misleading to a material extent, the Electoral Commissioner may request a person who has authorised, caused or published the matter or thing, to do one or more of the following:
   (a) withdraw the matter or thing from further publication;
   (b) publish a retraction in specified terms and in a specified manner and form.

(5) In deciding the terms, manner and form of a retraction requested under section 329A(4), the Electoral Commissioner must consider:
   (a) the terms, manner and form of the matter or thing published; and
   (b) the number of times the matter or thing had been viewed.

(6) If the Court is satisfied, on the balance of probabilities, that a matter or thing has been published in relation to electoral matters that is inaccurate or misleading to a material extent, the Court may order any person who has authorised, caused or published the matter or thing, to do one or more of the following:
   (a) withdraw the matter or thing from further publication;
   (b) publish a retraction in specified terms and a specified manner and form.

(7) Where a person consents to comply to a request under subsection (4) or an order is made under subsection (5) the Electoral Commissioner must publish a notice, in the manner prescribed by the regulations, notifying internet service providers and internet content hosts that such an order has been made or a request consented to.
   Note: A person can consent to comply with a request on a without-admissions basis.

(8) The Electoral Commissioner may make regulations for the purpose of establishing a notification scheme where members of the public or candidates may draw the Electoral Commissioner's attention to a purported offence under this section.

**Section 329B Obligations of internet service providers and internet content hosts relating to altered images**

(1) A person commits an offence if the person:
   (a) is an internet service provider or an internet content host; and
   (b) is aware that the service provided by the person can be used to access material that has been subject to an order or request under section 329A; and
   (c) does not refer details of the material to the Electoral Commissioner within a reasonable time after becoming aware of the existence of the material; or
   (d) if requested by the Electoral Commissioner, does not take reasonable steps to take down or remove access to that material.

(2) A person is presumed to be aware that a service they provide can be used to access material subject to an order or request under section 329A, where a

notice under section 329A(7) has been published and a reasonable period of time has elapsed.

(3) The Electoral Commissioner may provide guidance to organisations relating to their obligations under subsection (1). Any such guidance must be published in the manner prescribed in the regulations.

## VI CONCLUSION

Regulations that limit free speech must be suitable, necessary, and adequate in their balance. This article has considered the current protections available to politicians, political parties and the AEC to combat the growing threat posed by deepfake technology to elections, and by extension to democracy. It concludes that there are current gaps in the law, with copyright, tort and electoral law only offering very limited protections that could be readily avoided. These protections remain unclear, ill-defined and are inadequate to prevent the use of deepfakes to directly sway voter preferences, or to undercut truth in political discourse. In response, it proposes two targeted amendments to the *Electoral Act*. The amendments are, critically, both tailored proportionate to the threat posed by deepfakes. This article concludes that these measures are distinguishable from (appropriately rejected) calls for general regulations concerning misinformation or disinformation. While it would likely be possible to craft such laws, they would overburden free speech in Australia and lead to a significant chilling effect for media organisations, internet content platforms and everyday citizens, and reduce the strength of our democratic institutions.