THE UNIVERSITY OF MELBOURNE

Centre for Artificial Intelligence and Digital Ethics

CAIDE AI Policy Forums

FORUM #4 ISSUES PAPER: Demystifying Generative AI, Synthetic Content and the Pursuit of Truth

About the Author

Andrew Lim is a Research Associate with the Centre for Artificial Intelligence and Digital Ethics (CAIDE). He has just completed his Bachelor of Science in Physics and Diploma of Languages in Latin at the University of Melbourne. With prior experience across student advocacy and journalism, including guest lecturing on generative AI in a 'post-truth' age to young journalists at Boston College in the USA, Andrew is deeply passionate about the growing nexus between emerging technologies, public communications, and good governance in a globalised world: never more apparent than in questions of institutional trust, ground truth, and effective AI policy. Andrew would like to acknowledge the contributions of Professor Jeannie Marie Paterson to the editing of this paper.

What is CAIDE?

The Centre for Artificial Intelligence and Digital Ethics (CAIDE) is a cross-disciplinary research centre at the University of Melbourne. CAIDE facilitates cross-disciplinary research, teaching and leadership on the ethical, technical, regulatory and legal issues relating to AI and digital technologies. CAIDE is directed by Professor Jeannie Marie Paterson from Melbourne Law School. For more information about CAIDE, see our website: https://www.unimelb.edu.au/caide.

Acknowledgements and Sponsorships

The AI Policy forums are supported by:

- The Centre for AI and Digital Ethics, funded by the Faculty of Engineering and Information Technology and Melbourne Law School at the University of Melbourne
- The Ninian Stephen Law Program powered by the Menzies Foundation, as part of the project *New Legal Thinking for Emerging Technologies*
- Microsoft, Atlassian and the Tech Council of Australia as part of the project *Demystifying Generative AI*.

CAIDE AI Policy Forums

This issues paper was prepared as reading for the CAIDE AI Policy Forum #4, held at the University of Melbourne in 2024. It should be read in conjunction with the discussion paper on GenAI, synthetic content and truth, available here. The aim of the AI Policy Forums is to provide an opportunity for discussing policy and law issues raised by the emergence of AI that go beyond the headlines. This event considered the role of AI, particularly Generative AI, in the creation of synthetic content, and what that means for the function of truth in society.

Contact CAIDE

Email: uom-caide@unimelb.edu.au.
Website: https://www.unimelb.edu.au/caide.
LinkedIn: https://au.linkedin.com/company/caide-unimelb.

Demystifying Generative AI, Synthetic Content and the Pursuit of Truth

Background

The rise of easily accessible tools for generating synthetic content have raised serious concerns about their potential for societal and individual harm. Their use, especially in the form of political disinformation and intimate image abuse, have prompted some to call for regulatory action to ensure responsible usage of this technology. In this paper we focus on the response to misinformation and disinformation via synthetic consent, without undermining the importance of reforms on intimate image abuse and other harmful content online.

The Tech

Of particular concern have been so-called 'deepfake' images. A deepfake is strictly, 'a digital photo, video or sound file of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something they did not actually do or say'.[1] The concept is also sometimes used to describe entirely synthetic content generated by AI.[2] Most concern has been raised about deepfake videos, but increasingly synthesised voices are causing concern due to an inability to use visual synchronicity to identify deepfakes.[3]

Deep fakes can be entirely legitimate, such as in film and television. Deepfakes also have the potential for considerable individual, economic and societal harm.[4] Deep fakes and synthetic images may be used for intimate image abuse and extortion. They figure heavily in scams. In this paper, we focus primarily on political misinformation.

Figure 1: Indian Prime Minister Narendra Modi's face being mapped by an AI Media company to translate his speech into other languages for greater voter outreach in an example of 'softfake' usage. Himanshu Sharma/picture alliance via Getty Images

Other issues have been raised about language here – in particular, about the legitimacy of the word 'deepfakes', given its pejorative connotations and its origins in non-consensual pornographic material.[5] Some prefer the term 'softfake' to describe positive uses of this tech, like candidates in the 2024 Indian General Elections automatically translating video messages into a multilingual format. Others still point to the neutral term 'synthetic content', though this may be too broad, possibly referring to any content generated by AI. This issue does not necessarily exist outside English – in particular the Chinese term 换脸 ('face changing') carries more neutral ideas.

Political and Democratic Harms

Politically motivated deepfakes aimed at swaying voters ahead of elections have been observed to date in the United States[6], South Korea[7], Australia[8], India[9], Pakistan[10] and more. These span a wide variety of topics, quality and authors – both official political parties and individual actors alike.[11] Generating these deepfakes has also become significantly easier in recent years, thanks to very few safeguards being implemented on publicly available deepfake software, and those that exist being trivial to bypass with the use of other programs.[12] A May 2024 study by the Centre for Countering Digital Hate found that, of 240 attempts using 6 of the most popular voice cloning tools, in 80% of cases, convincing disinformation clones could be created of high-profile politicians, with one tool even auto-generating further speeches of disinformation from sparse prompts.[13]

More generally, the potential for disinformation in this technology has been highlighted by generative AI's use to sow doubt about the veracity of images in general (what Chesney and Citron dub "the liar's dividend"[14]). This phenomenon has been witnessed most clearly in the claiming real photos to have been ""A.I.'d"… [when] THEY DIDN'T EXIST! [sic]".[15] That said, the liar's dividend pays even when AI does not generate images that could be genuinely confused for real ones. In both the recent UK Southport riots[16] and the US presidential election campaign[17], obviously fake or cartoonish AI-generated images have been used to accentuate a line of attack more swiftly and at larger scales than before. These

have had very real consequences, with images attributed by media outlets to the surge of (baseless) claims about immigrants eating household pets in Springfield, Ohio that ultimately led to bomb threats in that city.[18]

The Proposals for Law Reform

With these potential harms, so too have emerged legislative, regulatory and technological proposals for solutions. Legal approaches have both taken the form of both catch-all deepfake legislation[19] and specific measures designed to deal with uses of deepfakes like political misinformation. They have also ranged from outright bans and disclaimer requirements to more technologically focussed solutions like digital fingerprinting and watermark requirements. Regulators like the FCC in the United States have also utilised existing powers to crack down on actions where possible (for instance, using telephone consumer law provisions against voter-misinformation deepfake robocalls).[20] And technology-based detection software systems also purport to provide an effective defence against such content.[21]

Figure 2: Even synthetic content not confusable for truth can help spread disinformation, with some arguing broader democratic harms remain. Donald Trump/Truth Social

The truth is that no one piece of legislation or technological advancement will, however developed, provide an easy solution. Rather, each one presents its own unique set of trade-offs between competing desires – for easy-to-navigate regulatory frameworks; for the retention of free and fair elections; and much more. Any policy solution must consider both normative legal concerns and practical technical realities, reflecting the breadth of these issues and stakeholders to craft an integrated policy approach.

The Policy Issues

1: Democratic Speech – a special case?

Given the widespread use of deepfakes in the form of election manipulation and 2024's status as a 'year of elections' worldwide[22], it is telling that some of the most developed legislative solutions globally have been targeted at the use of deepfakes and other AI tools for disinformation and election manipulation.

This has been most pronounced in the United States, where at both a federal and state level, a plethora of bills have been introduced. Some, like the Protect Elections from Deceptive AI Act[23] seek explicitly to ban AI-generated deepfake material, whereas others, like the AI Transparency in Elections Act[24] seek only to require disclaimers. Little, however, has passed federally (despite bipartisan co-sponsorships on these and other bills). At a state level, there is much variance across the country, many involving the creation of new criminal offences related to political deepfakes.[25] South Korea has had more success, revising the Public Official Election Act[26] in 2023 to ban deepfakes during the 90-day campaign period.

In Australia, this discussion has been broadly dominated by the possibility of truth in political advertising legislation, seeking to create new offences, new regulators or extend the misleading and deceptive publications offence in s 329 of the *Commonwealth Electoral Act 1918* (Cth) to include facts that would sway election decisions.

The Australian government signalled in March this year that it would be introducing such legislation by mid-2024[27], and, following a period of government inactivity where pressure for these measures largely came from the crossbench,[28] the work was folded into the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024 (Cth), introduced in September. Unlike its draft bill predecessor, the bill as introduced contained no exemptions for government and political party content. It further empowers the Australian Communications and Media Authority (ACMA) to develop binding standards for digital communications platform providers on misinformation and

disinformation where they can cause 'serious harm'. ACMA would be able to seek civil penalties up to 5% of annual turnover from social media companies who host such content.

However, following substantial opposition from the Greens, the Coalition and several Senate crossbenchers from both free speech and enforceability concerns,[29] the government determined in November that "no pathway to legislate this proposal" remained, withdrawing the bill.[30] Announcing this, the government highlighted other efforts to deal with synthetic content. These include the *Criminal Code Amendment (Deepfake Sexual Material) Act 2024* (Cth), election truth in political advertising reforms (which appear to have stalled again, despite Special Minister of State Don Farrell's intent to legislate this before the next election and reported briefing of crossbenchers to that effect[31]) and wider AI regulation from the Minister for Industry and Science.

In any event, while broad notions of truth in political advertising legislation may hold public support[32], concerns remain over the constitutionality of actual such legislation and the role of the *Australian Constitution*'s implied freedom of political communication.[33] Indeed, more broadly, it has been free speech provisions that have proven problematic for overseas legislation regulating political advertising, with the UK exempting political advertising from its non-broadcast standards[34] and the US legislation being delayed in part due to concerns over free speech.[35] Particularly in the latter jurisdiction, this concern is not limited to political misinformation focussed legislation, but has also affected more general deepfake bills holding individuals liable for unauthorized digital replicas[36], though not those concerned with watermarking or digital fingerprinting techniques.[37]

2: The Case for Transparency

A key element of regulatory reform proposals in many jurisdictions has focused on transparency for all artificially generated content. This is a key element of the new EU AI Act. Article 50 provides[38]:

> 2. Providers of AI systems, including general-purpose AI systems, generating synthetic audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Providers shall ensure their technical solutions are effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards. This obligation shall not apply to the extent the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof, or where authorised by law to detect, prevent, investigate or prosecute criminal offences.

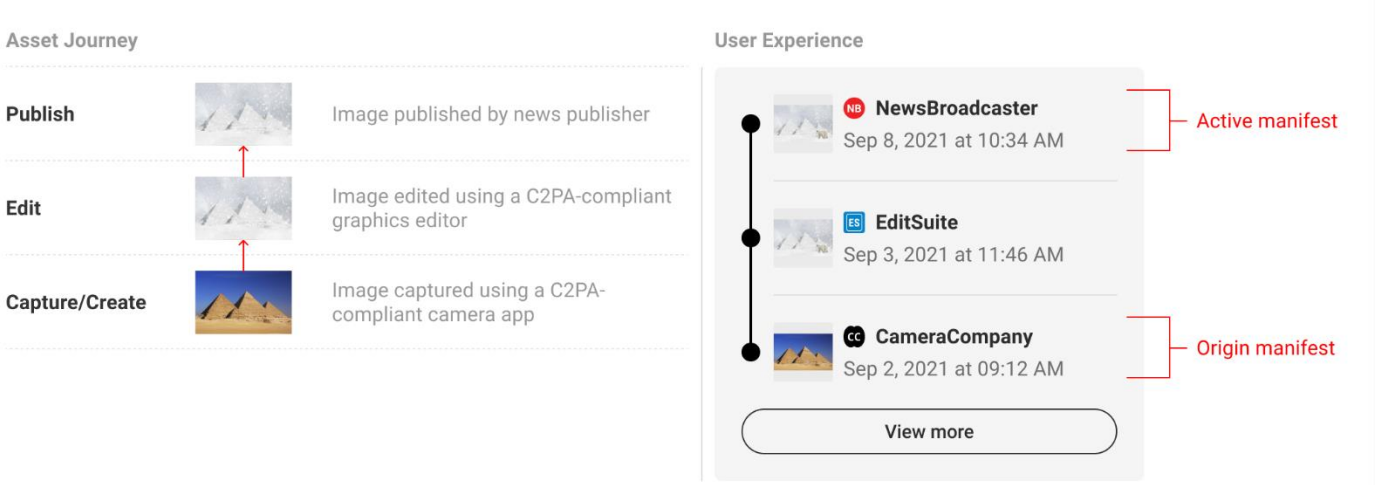

Figure 3: The C2PA technical standard is one technical solution to transparency issues – giving users a timeline of modifications to an image/video/audio attributed to trusted entities and programs, confirming its provenance (though not its truthfulness). C2PA

In Australia, the Department of Industry, Science and Resources outlined possible approaches for mandatory guardrails for AI in its September 2024 *Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings*.[39] The paper contemplates a risk-based approach that differentiates low-, medium- and high-risk use cases, emphasising in

high-risk scenarios the implementation of *ex ante* measures to prevent misuse rather than create liability (and induce expensive litigation) through a series of mandatory guardrails.[40]

Of special note to the deepfake discussion is Guardrail 6, the duty to "inform end users regarding AI-enabled decisions, interactions with AI and AI-generated content", which notably includes the requirement that "Organisations must apply best efforts to ensure AI-generated outputs, including synthetic text, image, audio or video content, can be detected as artificially generated or manipulated."[41]

The discussion paper also raises different models for embedding the regulatory regime. This raises the question of which regulators would be responsible for deep fake's affecting political and social discourse. The AEC, for instance, might be empowered to deal with deepfakes in authorised electoral communications; ACMA to consider social media more broadly – and possibly decisions would be cross-referenced to a national 'AI' regulator responding to identified categories of risk.

In the meantime, piecemeal legislation addressing individual deepfake and disinformation concerns will continue to emerge. These include not just the misinformation and disinformation measures mentioned in Section 1, but also other attempts to mitigate online harms, whether by placing age limits on social media[42] or reforming the *Online Safety Act* 2021 (Cth) to better reflect the generative AI landscape.[43]

3: Strengthening Institutions– how much new statute law is really needed?

All this said, there may also be a place for existing legislation and institutions, especially in Australia, to provide the protections necessary to deal with deepfake-related issues. In general, a statute-light approach has been seen by some as not only inevitable but desirable, placing AI in a lineage of other technological developments of the past few centuries that did not require technology-specific legislation for effective regulation. As a Productivity Commission report in January 2024 put it:

> "Many potential harms have been encountered with past technologies and adequately dealt with by existing regulatory frameworks in areas such as consumer protection, privacy, anti-discrimination, negligence and sector-specific and profession-specific requirements. AI is no different"[44]

However, this approach has also raised questions about the role of *ex ante* regulation in general: given the Draghi report's stark warnings of a lack of technological competitiveness and economic burdens in the EU created by prohibitive legal guardrails[45], there are fears Australia too is heading down a route of over-regulation in digital spaces more generally.

This view is welcomed by some due to the high cost inherent in a principles-based legislative approach. For every slight tweak in a system, a whole process of recategorization, risk assessment and harm mitigation would be necessary under a strict transparency approach: the cost of this may be so prohibitive it drives startups and smaller companies (unable to simply absorb these costs) out of Australia into jurisdictions with less stringent requirements or higher populations of potential users (like the US).

Global AI governance has also been floated as a potential solution, with views towards internationally recognised governing bodies. That said, with this technology still in its infancy, it is notable that most unified international regulatory regimes have taken decades to develop, so any such development is likely far off.

4: Technical or Social: which is the Post-Truth Problem?

Relying on existing regulation is not the only alternative solution to AI-specific law. We may recall Easterbrook's 1990s cyberlaw invocation of the 'Law of the Horse': even if many laws concern one subject – say, horse sales, horse veterinary care, and horse-induced injuries –their defining characteristic is not necessarily 'horse law', nor is such a grouping necessarily meaningful and/or useful.[46] In a similar vein, some fear the rush to legislate for AI has incorrectly grouped substantively different concerns into this catch-all legislation. After all, worries about a post-truth world predate the generative AI boom[47] – and the liar's dividend also pays for non-deepfake-related denials of fact.[48]

To some scholars, the wider normative concern is neither regulatory nor technological but about institutional trust – the product of a decades-long decline in trusted local news sources that has diminished the role of an active and trusted

news media in fact-checking, allowing the liar's dividend to take hold.[49] Simon, Altay and Mercier even suggest that generative AI's threat is ultimately "rare in the information environment of wealthy, democratic countries….thanks to the hard work of…journalists, fact checkers [and] experts."[50]

This is particularly concerning for solutions like C2PA (see Figure 3), a technical standard aimed at ensuring provenance of content can be easily determined.[51] While such solutions address the technical issue, they can just as easily be ignored by individual, as is their democratic right. This is not to say that no problem exists but rather that that problem is one of social policy: about an apathetic public rather than a lagging legal system.

In particular, the decisions to increasingly reduce teaching of civics, politics and history in the Australian Curriculum are of significant concern.[52] While compulsory voting may induce forced engagement in politics, it does not necessarily induce active and critical thinking around it. A lack of trust in institutions like journalism and the loss of local news sources may prove dangerous for the 'truth' irrespective of specific generative AI concerns, and bolstering such institutions, whether via public funding or via private sector initiatives like Microsoft's Democracy Forward project,[53] may prove a more effective solution.

Measures often suggested in response include strengthening these institutions through media literacy programs[54], expansions of reimbursement schemes like Australia's News Media Bargaining Code[55], or even exacting a media levy to directly seek payment from social media platforms and ensure journalism funding.[56]

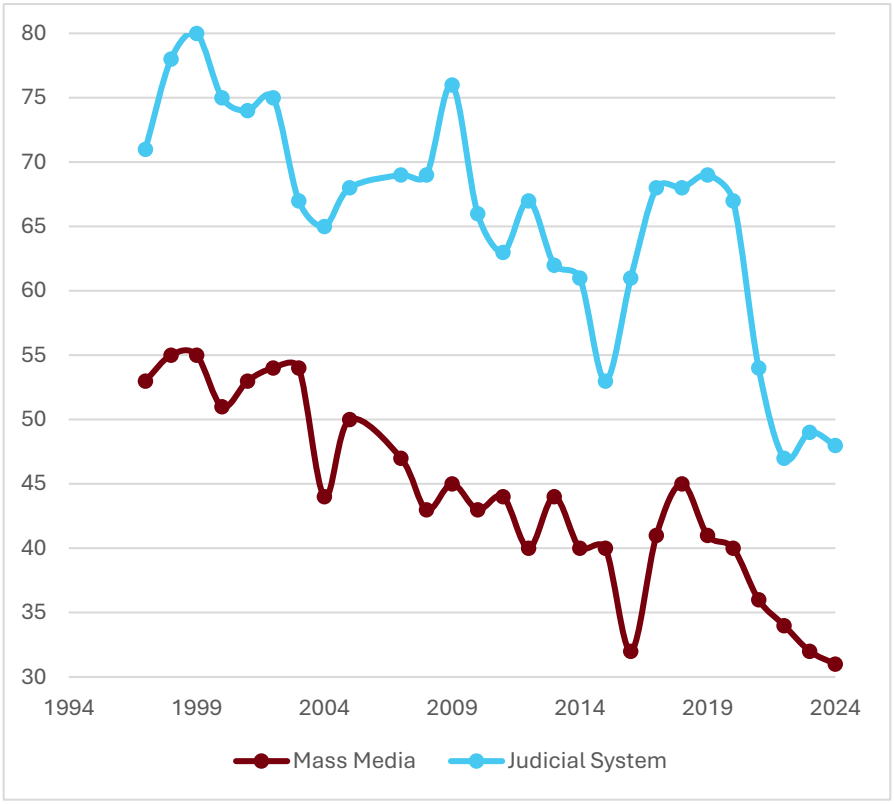

Figure 4: Percentage of Americans who have trust and confidence in the mass media and the judicial system, showing a distinct recent decline. Some argue trends like these, seen across the globe, prove the real normative concern in a post-truth world. Data: Gallup / Visuals: CAIDE

Figure 1: Table of Proposed Australian Measures on GenAI and Truth (Updated 29 November 2024)

Measure	Status	Key Measures
Digital Duty of Care	Announced after recommendation in *Online Safety Act* statutory review	Places obligation on tech companies to take reasonable steps to identify, mitigate and prevent potential risks and harms that fall into certain legislated categories.
Online Safety Act 2021 (Cth) Basic Online Safety Expectations (BOSE)	BOSE amended in June 2024	Requires reports on child/adult user breakdown, includes 'best interests of the child' test, and emphasises user safety criteria for generative AI
Online Safety Amendment (Social Media Minimum Age) Bill 2024 (Cth)	Passed; Awaiting Royal Assent	Sets a minimum age of 16 years for access to social media platforms, and introduces an obligation on providers to take 'reasonable steps' to use age assurance to prevent underage users from accessing social media, with failure to do so resulting in fines up to $49.5 million. Excludes messaging apps, education/health services and online gaming services from the definition of social media platforms.
Digital Platform Levy	Recommended by Select Committee	Levy exacted on digital platforms to extract funding for public interest journalism aimed at strengthening institutional protections against mis/dis-information.
Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024 (Cth)	Withdrawn due to Senate opposition Further post-election attempts possible	Requires providers to report on misinformation risk on platforms. Empowers ACMA to create misinformation and disinformation standards, codes and dispute resolution procedures, with civil penalties up to 5% of annual turnover for social media companies. Includes explicit carve-outs for mainstream media/news, academic, artistic, satirical, humorous, scientific and religious speech

[1] eSafety Commissioner, 'Deepfake trends and challenges – position statement' (19 August 2024) <https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes>.
[2] Hannah Smith and Katherine Manstead, *Weaponised Deep Fakes – National Security and Democracy* (ASPI Policy Brief, 29 April 2020) <https://www.aspi.org.au/report/weaponised-deep-fakes>.
[3] Roman H Kepczyk, 'Deepfakes emerge as real cybersecurity threat,' *AICPA-CIMA* (28 September 2022) <https://www.aicpa-cima.com/news/article/deepfakes-emerge-as-real-cybersecurity-threat>
[4] Bobby Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 *California Law Review* 1753.
[5] See generally, Samantha Cole, 'Deepfakes Were Created As a Way to Own Women's Bodies—We Can't Forget That,' *VICE* (18 June 2018) <https://www.vice.com/en/article/deepfake-porn-origins-sexism-reddit-v25n2/>
[6] In both the upcoming 2024 Presidential Election and for the Massachusetts Senate election – see, among others (despite names, all originally spread without disclosure that they weren't factual) Mr Reagan, 'Kamala Harris Ad PARODY 2' (YouTube, 31 July 2024) <https://www.youtube.com/watch?v=RIjoPqIcyaw>; Rob Lever, 'Deepfake video of Elizabeth Warren spreads on TikTok', *AFP Fact Check* (30 April 2024) <https://factcheck.afp.com/doc.afp.com.33AE6KT> FRANCE 24 English, 'No, Taylor Swift did not endorse Donald Trump at the Grammy Awards • FRANCE 24 English' (YouTube, 7 February 2024) <https://www.youtube.com/watch?v=2JPv4rTS-KU>.
[7] In the 2024 Parliamentary Elections – see Editorial, 'Deepfake risks in election', *The Korea Herald* (Seoul, 23 February 2024) <https://www.koreaherald.com/view.php?ud=20240222050761> and 'KCSC Approves Police Request to Delete, Block Fabricated Content on Pres. Yoo', *KBS World* (23 February 2024) <https://world.kbs.co.kr/service/news_view.htm?lang=e&Seq_Code=183867>.
[8] For the upcoming Queensland State Election – Andrew Messenger, 'Queensland premier rules out AI-generated election material after LNP releases dancing TikTok attack advertisement', *The Guardian* (23 July 2024) <https://www.theguardian.com/australia-news/article/2024/jul/23/queensland-premier-steven-miles-ai-generated-qld-election-video-lnp-tiktok-attack-ad>.
[9] Nilesh Christopher, 'How AI is resurrecting dead Indian politicians as election looms', *Al Jazeera* (12 February 2024) <https://www.aljazeera.com/economy/2024/2/12/how-ai-is-used-to-resurrect-dead-indian-politicians-as-elections-loom>
[10] Gibran Peshimam, 'Pakistan's jailed ex-PM Imran Khan claims election victory', *Reuters* (10 February 2024) <https://www.reuters.com/world/asia-pacific/pakistans-jailed-ex-pm-imran-khan-claims-election-victory-2024-02-09/>.
[11] For a broader (if rather Americentric) database, see the still-updated database introduced in Christina P Walker, Daniel S Schiff, and Kaylyn Jackson Schiff, 'Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database' (2024) 38(21) *Proceedings of the AAAI Conference on Artificial Intelligence* 23053 <https://doi.org/10.1609/aaai.v38i21.30349>, available at http://bit.ly/pdid.
[12] Centre for Countering Digital Hate, *Attack of the Voice Clones: How AI voice cloning tools threaten election integrity and democracy* (31 May 2024), 9-10. <https://counterhate.com/wp-content/uploads/2024/05/240524-Attack-of-the-Voice-Clones-REPORT_final.pdf>
[13] Ibid, 9-12.
[14] Bobby Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 *California Law Review* 1753, 1785-6.
[15] @realDonaldTrump (Truth Social, 12 August 2024, 12:09pm EDT) <https://truthsocial.com/@realDonaldTrump/posts/112944255426268462>
[16] Ben Quinn and Dan Milmo, 'How TikTok bots and AI have powered a resurgence in UK far-right violence', *The Guardian* (2 August 2024). <https://www.theguardian.com/politics/article/2024/aug/02/how-tiktok-bots-and-ai-have-powered-a-resurgence-in-uk-far-right-violence>

[17] See, e.g., @elonmusk (X, 3 September 2024, 1:18pm EDT) <https://x.com/elonmusk/status/1830656672211103825?s=46&t=SC7A70vobuZXZOd40GdsKg>.
[18] David Ingram, 'How AI images of cats and ducks powered the pet-eating rumor mill in Springfield, Ohio', *NBC News* (15 September 2024) <https://www.nbcnews.com/tech/misinformation/ai-images-cats-ducks-powered-pet-eating-rumor-mill-rcna171065>.
[19] See, e.g., in the United States, the Nurture Originals, Foster Art, and Keep Entertainment Safe (No FAKES) Act, S 4875, 118th Congress (2024) and the No Artificial Intelligence Fake Replicas And Unauthorized Duplications (No AI FRAUD) Act, HR 6943, 118th Congress (2024).
[20] See Federal Communications Commission (US), 'FCC Makes AI-Generated Voices in Robocalls Illegal' (News Release, 8 February 2024) <https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal> and Federal Communications Commission (US), 'FCC-State Robocall Investigation Partnerships' (11 March 2024) <https://www.fcc.gov/fcc-state-robocall-investigation-partnerships>.
[21] See Cartier et al (n 19) and Lenovo Storyhub, 'McAfee Launches World's First Automatic and AI-powered Deepfake Detector Exclusively on Select New Lenovo AI PCs' (Press Release, 21 August 2024) <https://news.lenovo.com/pressroom/press-releases/mcafee-first-automatic-ai-powered-deepfake-detector/>.
[22] See, e.g., Toby James and Holly Ann Garnett, 'Half the world will vote in 2024, but how many elections will be fair?', *The Conversation* (16 March 2024) <https://theconversation.com/half-the-world-will-vote-in-2024-but-how-many-elections-will-be-fair-225828>.
[23] Protect Elections from Deceptive AI Act, S 2770, 118th Congress (2024).
[24] AI Transparency in Elections Act, S 3875, 118th Congress (2024).
[25] See, among others, the in-force Tex Election Code §255.004 (2019); Act of 16 May 2024, 2024 Ala Laws 349; Ariz Rev Stat Ann §16-1023 (2024) and Cal Elections Code §20010 (2022). See also Defending Democracy from Deepfake Deception Act, Cal AB2655 (2024).
[26] Seungmin Lee, 'AI and Elections: Lessons from South Korea', *The Diplomat* (13 May 2024). <https://thediplomat.com/2024/05/ai-and-elections-lessons-from-south-korea/>
[27] Paul Karp, 'Peter Dutton says truth in political advertising 'probably welcome' but criticises Labor as scare campaign 'experts'', *The Guardian* (14 March 2024) <https://www.theguardian.com/australia-news/2024/mar/14/peter-dutton-truth-in-political-advertising-laws-labor-policy>
[28] Jordyn Beazley, 'David Pocock calls for election ban on AI deepfakes with fake videos of Albanese and Dutton', *The Guardian* (7 September 2024) <https://www.theguardian.com/technology/article/2024/sep/07/david-pocock-deepfake-video-anthony-albanese-peter-dutton-ai-election-regulation>
[29] Josh Butler, 'Labor dumps misinformation bill after Senate unites against it', *The Guardian* (24 November 2024) <https://www.theguardian.com/australia-news/2024/nov/24/labor-dumps-misinformation-bill-after-senate-unites-against-it>.
[30] Michelle Rowland (Minister for Communications), 'Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024' (Media Release, 24 November 2024) <https://minister.infrastructure.gov.au/rowland/media-release/communications-legislation-amendment-combatting-misinformation-and-disinformation-bill-2024>.
[31] Matt Martino, 'Deepfakes and falsehoods are legal in political advertising. Not everyone is on board with fixing it', *ABC News* (15 October 2024) <https://www.abc.net.au/news/2024-10-15/deepfakes-misinformation-ai-gen-in-political-advertising-legal/104470006>.
[32] The Australia Institute, 'Overwhelming support for truth in political advertising laws following referendum' (Media Release,19 October 2023) <https://australiainstitute.org.au/post/overwhelming-support-for-truth-in-political-advertising-laws-following-referendum/>.
[33] Kieran Pender, 'Regulating Truth and Lies in Political Advertising: Implied Freedom Considerations' (2022) 44(1) *Sydney Law Review* 1.
[34] Nick Harrison, 'Online political advertising in the UK | Democracy in the Digital Age', *Taylor Wessing* <https://www.taylorwessing.com/en/interface/2024/democracy-in-the-digital-age/online-political-advertising-in-the-uk>
[35] See, e.g., the discussion resulting in Sen Fischer's objection to unanimous consent motions on several deepfake election regulation bills at 170 *Congressional Record* S5654-6 (daily ed, 31 July 2024).
[36] See, on the US' No FAKES Act discussed earlier, Katherine Klosek, 'No Frauds, No Fakes…No Fair Use?', *Association of Research Libraries (ARL) Views* (19 April 2024) <https://www.arl.org/blog/nofraudsnofakes/>
[37] See, from the US, the Protecting Consumers from Deceptive AI Act, HR 7766, 118th Congress (2024).
[38] *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828* [2024] OJ L 2024/1689, art 50.
[39] Department of Industry, Science and Resources (Cth), *Proposals Paper for Introducing Mandatory Guardrails for AI in High-Risk Settings* (September 2024). <https://storage.googleapis.com/converlens-au-industry/industry/p/prj2f6f02ebfe6a8190c7bdc/page/proposals_paper_for_introducing_mandatory_guardrails_for_ai_in_high_risk_settings.pdf>
[40] Ibid 16-17.
[41] Ibid 39.
[42] Prime Minister of Australia, 'Albanese Government set to introduce minimum age for social media access' (Media Release, 10 September 2024) <https://www.pm.gov.au/media/albanese-government-set-introduce-minimum-age-social-media-access>.
[43] Michelle Rowland (Minister for Communications), 'Government to introduce legislation to combat seriously harmful misinformation and disinformation' (Media Release, 12 September 2024) <https://minister.infrastructure.gov.au/rowland/media-release/online-safety-expectations-boost-transparency-and-accountability-digital-platforms>.
[44] 'Making the most of the AI opportunity: The challenges of regulating AI' (Research Paper No 2, Productivity Commission (Cth), 2024) 1. <https://www.pc.gov.au/research/completed/making-the-most-of-the-ai-opportunity/ai-paper2-regulating.pdf>
[45] Mario Draghi, *The future of European competitiveness: Part A | A competitiveness strategy for Europe* (9 September 2024) <https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en>
[46] Frank H Easterbrook, 'Cyberspace and the Law of the Horse' [1996] *University of Chicago Legal Forum* 207 <https://chicagounbound.uchicago.edu/uclf/vol1996/iss1/7/>.
[47] See, for a useful and pre-widespread-generative-AI view on the subject, Lee McIntyre, *Post-Truth* (The MIT Press, 2018).
[48] Indeed, a recent study suggests this is in fact more effective than the deepfake liar's dividend. See Kaylyn Jackson Schiff, Daniel S Schiff and Natália S Bueno, 'The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?' [2024] *American Political Science Review* 1. <https://www.cambridge.org/core/journals/american-political-science-review/article/liars-dividend-can-politicians-claim-misinformation-to-evade-accountability/687FEE54DBD7ED0C96D72B26606AA073>
[49] See Margaret Sullivan, *Ghosting the News: Local Journalism and the Crisis of American Democracy* (Columbia Global Reports, 2020).
[50] Felix M Simon, Sacha Altay and Hugo Mercier, 'Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown' (2023) 4(5) *Harvard Kennedy School (HKS) Misinformation Review* <https://misinforeview.hks.harvard.edu/wp-content/uploads/2023/10/simon_generative_AI_fears_20231018.pdf>.
[51] Coalition for Content Provenance and Authenticity, 'Overview – C2PA' <https://c2pa.org/>
[52] Peter Brett, 'The insidious way the new curriculum undermines democracy,' *Australian Association for Research in Education* (16 May 2022). <https://blog.aare.edu.au/the-insidious-way-the-new-curriculum-undermines-democracy/>

[53] Microsoft, 'Democracy Forward | Microsoft Corporate Social Responsibility' <https://www.microsoft.com/en-us/corporate-responsibility/democracy-forward>

[54] See, in the US, the Artificial Intelligence Literacy Act, HR 6791, 118th Congress (2023). Similar state legislation discussed in Justin Klawans, 'The push for media literacy in education amid the rise of AI', *The Week* (2 April 2024) <https://theweek.com/tech/media-literacy-AI-schools>.

[55] See, among others, Anya Schiffrin, *AI and the future of journalism: an issue brief for stakeholders*, UN Doc CI-2024/WTR/3 (2024) <https://unesdoc.unesco.org/ark:/48223/pf0000391214> and Sam Buckingham-Jones, 'Use news code, not copyright, to make AI firms pay for content: Sims', *The Australian Financial Review* (22 June 2023) <https://www-afr-com.eu1.proxy.openathens.net/companies/media-and-marketing/use-news-code-not-copyright-to-make-ai-firms-pay-for-content-sims-20230622-p5dijz>.

[56] Ben Doherty, 'Meta and Google face 'big tech tax' as Labor plans how to ensure media sustainability in Australia', *The Guardian* (21 October 2024). <https://www.theguardian.com/australia-news/2024/oct/21/meta-and-google-face-big-tech-tax-as-labor-plans-how-to-ensure-media-sustainability-in-australia>

THE UNIVERSITY OF
MELBOURNE

Centre for AI and Digital Ethics

Level 8
Melbourne Connect
700 Swanston Street
Carlton 3053

unimelb.edu.au/caide