Australian Human Rights Commission

Celebrating 40 years

Breaking trust:

How synthetic media is undermining democracy and human rights in Australia

November 2026



humanrights.gov.au

Acknowledgements

Authors: Lorraine Finlay, Sallie Harrington Downie and Patrick Hooton.

The Australian Human Rights Commission thanks the following people for their contributions:

- President Hugh de Kretser
- Australian Human Rights Commission staff for their review and input including: Darren Dick, Sabrina Dong, Lauren Cooper, Clara van den Bosch and Sascha Wightman. The Commission's Internal Working Group comprising of Sabrina Dong, Romy Groth, Polly Jenner, Saranya Kudasamy, Ruby Mountford, Susan Newell, Susan Nicolson, Christine Ratnasingham, Lubna Sherieff and Sascha Wightman.
- Wotton Kearney staff for their research assistance including Leanne Ho, Finn McNamara, Leah Mooney, Claire Thurstans, Lillian Bender, Samantha Younane, and Ben Quick
- Stakeholders from around Australia whose engagement helped to inform this report.

Design and layout: Claudia Williams

Acknowledgement of Country

The Australian Human Rights Commission acknowledges the Traditional Custodians of Country throughout Australia, and recognises their continuing connection to land, waters and culture. We pay our respects to their Elders – past and present.

Copyright

© Australian Human Rights Commission 2026

All material presented in this publication is licensed under the Creative Commons Attribution 4.0 International Licence (CC BY 4.0), except for:

- photographs, images, logos, any branding or trademarks
- content or material provided by third parties
- where we say otherwise.

You can copy, communicate and adapt this publication, as long as you attribute the Australian Human Rights Commission and abide by the other licence terms.

Suggested citation: Australian Human Rights Commission (2026), *Breaking Trust: How Synthetic Media is undermining Democracy and Human Rights in Australia,* Australian Human Rights Commission, Sydney.

ISBN: 978-1-921449-33-8

About the Australian Human Rights Commission

Our vision for Australia is one where everyone can live free and equal in dignity and rights as part of a strong, healthy democratic society.

We have 4 core functions under our legislation:

Access to justice: We help people to resolve complaints of discrimination and human rights breaches through our complaint handling services.

Fairer laws, policies and practices: We review existing and proposed laws, policies and practices and provide expert advice on how they can better protect people's human rights. We help organisations to protect human rights in their work. We publish reports on human rights problems and how to fix them.

Education and understanding: We promote understanding, acceptance and public discussion of human rights. We deliver workplace and community human rights education and training.

Compliance: We are the regulator for the positive duty requiring employers and others to address workplace sexual harassment, sex discrimination, sexist conduct and related victimisation.

Contact us

Contact the Australian Human Rights Commission:

online: www.humanrights.gov.au

by mail: GPO Box 5218 Sydney NSW 2001

by phone: 02 9284 9600

Or find us on social media:

Facebook

Instagram

LinkedIn

YouTube



Breaking trust:

How synthetic media is undermining democracy and human rights in Australia

November 2026

Contents

Commissioner's foreword **5**

1. Executive summary **7**
1.1 Definitions 7
1.2 Impact on human rights 8
1.3 Impact on democracy 10
1.4 Responding to impacts 11
1.5 Growing scale and sophistication 12
1.6 Builing democratic resilience 13

2. Adverse impacts on human rights **16**
2.1 Women's right to participation and privacy 18
2.2 Equality and non-discrimination 21
2.3 Freedom of expression and manipulation 24
2.4 Interference with peaceful assembly 27
2.5 Manipulation of free thought 28
2.6 Access to reliable information 29

3. Democratic resilience **31**
3.1 Erosion of trust in institutions and figures 32
3.2 Interference in Australian democracy 40

4. A layered response to synthetic media **45**
4.1 Electoral laws 47
4.2 Deepfake sexual abuse material laws 51
4.3 Misinformation and disinformation laws 54
4.4 Other areas of law 55
4.5 Technical responses 60
4.6 Education 63
4.7 Independent news and media 65
4.8 Trusted institutions 65

5. Context **67**
5.1 Method 67
5.2 Consultation 67

6. Appendix **68**
Appendix A: Definitions 68

Commissioner's foreword

Human rights and democracy are deeply interconnected.

Human rights protect the dignity, freedom and equal participation of every person. They enable people to express their views, access information, participate in public affairs, engage in public debate and hold institutions to account. In doing so, they help create the conditions in which democratic societies can flourish.

Synthetic media is challenging those conditions.

Much of the public discussion about deepfakes has focused on the risk of deception. That concern is justified. Synthetic media can be used to falsely depict people saying or doing things they never said or did. It can be used to spread misinformation, manipulate public opinion and undermine democratic processes.

But one of the central findings of this report is that the harms of synthetic media extend beyond whether people actually believe the content itself.

Synthetic media can interfere with a range of human rights, including privacy, equality, freedom of expression, freedom of thought, access to information and participation in public affairs. It can cause significant harm to individuals whose image, voice or identity is misused without their consent. It can damage reputations, facilitate harassment and abuse, and discourage participation in public life. These harms are often experienced most acutely by women and marginalised communities, who already face disproportionate barriers to participation and representation.

At the same time, synthetic media can weaken democratic participation even when it is recognised as false. By creating uncertainty about what is authentic, enabling genuine evidence to be dismissed as fake, and undermining confidence in information, institutions and media, synthetic media can make informed democratic engagement more difficult. In this way, its effects are felt not only by those directly targeted, but by society more broadly.

This report approaches synthetic media through a human rights lens because the impacts of these technologies are ultimately experienced by people. While the technology is new, the values at stake are not. Questions about dignity, autonomy, participation, accountability and trust have long been central to both human rights protection and democratic governance.

The report also recognises that responding to these challenges requires careful judgment. Measures designed to address the harms associated with synthetic media must themselves respect human rights and democratic freedoms. Protecting individuals, communities and democratic institutions cannot come at the expense of freedom of expression or legitimate political communication. Equally, a commitment to free expression should not prevent action to address genuine harm. The challenge is not to choose between these objectives, but to balance them appropriately.

One of the clearest lessons from this work is that there is no single solution. Legal protections, technical safeguards, authenticity measures, digital and civics education, independent media, responsible content creation and trusted public institutions all have an important role to play. The most effective response will be one that recognises the strengths and limitations of each and brings them together as part of a coordinated and layered approach.

The issues examined in this report will continue to evolve. Synthetic media will become more sophisticated. New opportunities will emerge alongside new risks. Yet the fundamental question will remain the same: how do we ensure that technological change strengthens, rather than diminishes, dignity, human rights and democratic participation?

The report also recognises that responding to these challenges requires careful judgment. The legal and policy environment relating to emerging technologies is evolving. Governments are considering new regulatory approaches to address emerging digital risks and harms. Measures designed to address the harms associated with synthetic media must themselves respect human rights and democratic freedoms. Protecting individuals, communities and democratic institutions cannot come at the expense of freedom of expression or legitimate political communication. Equally, a commitment to free expression should not prevent action to address genuine harm. The challenge is not to choose between these objectives, but to balance them appropriately.



Lorraine Finlay

Human Rights Commissioner

1. Executive summary

Democracy and human rights are inseparable foundations of a free society. Democracy creates the conditions for rights to flourish, while human rights (e.g. freedom of expression, public participation) are necessary for democracy to function. If human rights are not protected, democracy becomes hollow.

This report examines how the growing creation and use of synthetic media has potential to undermine human rights and assesses the implications for democratic participation and resilience in Australia.

Synthetic media gives rise to two related but distinct forms of harm. The first occurs when a person's image, voice or identity is used to create false or manipulated content. This can cause direct harm to the individual, including by interfering with their privacy, dignity, equality, freedom of expression and ability to participate in public life.

The second arises when synthetic media makes it harder to distinguish real information or events from what is false. People may be led to believe events occurred when they did not, or to dismiss genuine information as fabricated. While this can harm individuals, its effects may also extend more broadly by weakening trust in information and democratic processes.

Around the world democracies are backsliding, and Australia is not immune.[1] This coincides with the rise of artificial intelligence (AI)[2] and social media platforms that have fundamentally changed the information environment.[3] Democracy advocates have raised concern about AI's ability to amplify existing threats to democracy.[4] The Grattan Institute has noted that AI has played a role in elections in at least 50 countries (including Australia).[5]

1.1 Definitions

There is no single, consistently accepted definition of a deepfake or synthetic media. The technology and the language used to describe it are also continuing to evolve.

Defining terms

For clarity, this report uses the following definitions:

- **Deepfake:** Synthetic media which is an AI-generated or AI-altered image, video or audio of a real person doing or saying something that they did not actually do or say.[6]
- **Synthetic media:** an umbrella term for any AI-generated or altered photo, video or audio (including deepfakes).

This report uses synthetic media as a broader term because some of the problems the report examines do not involve the imitation of a particular person. AI-generated content may instead fabricate an event or falsely depict conduct by an institution or group. These forms of content can affect human rights and democratic participation even if they do not fall within a narrower definition of a deepfake.

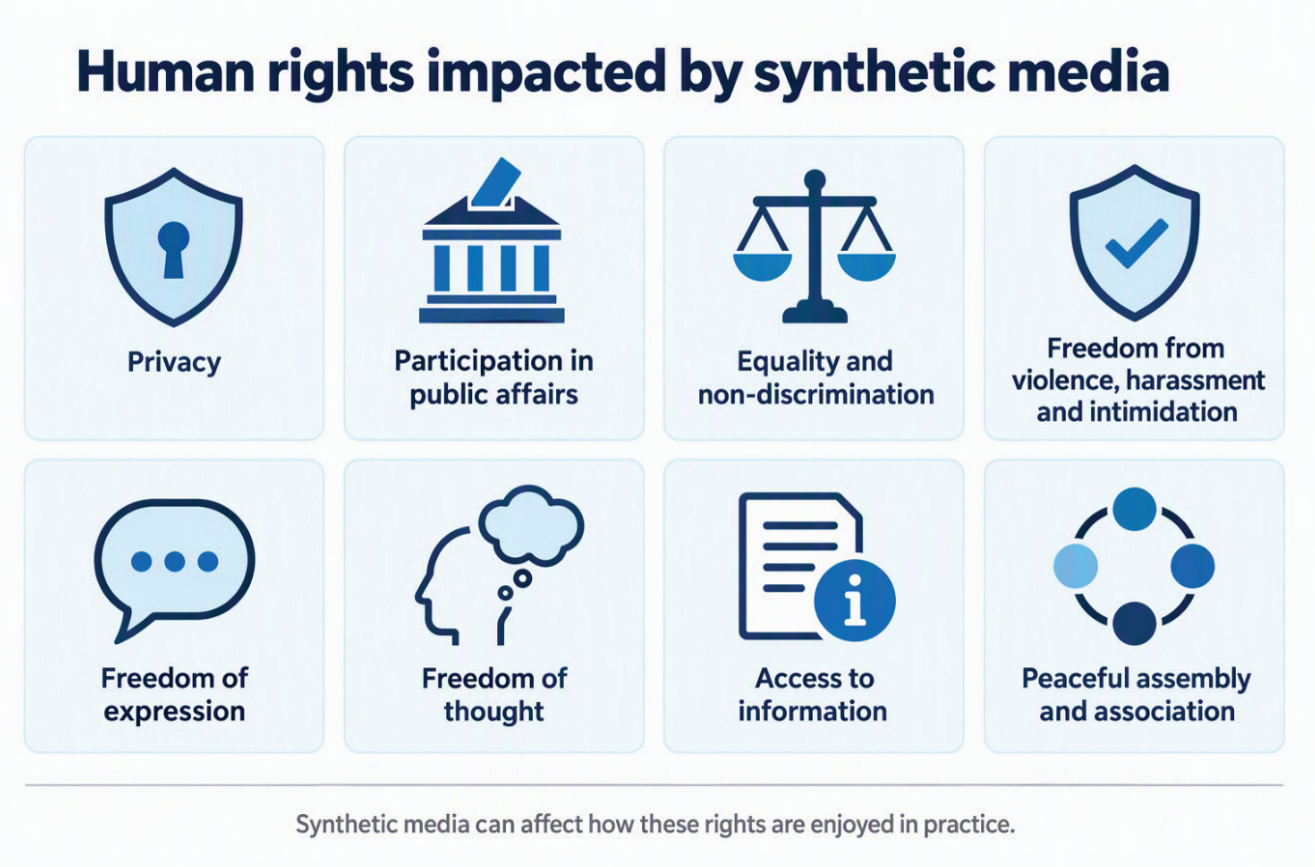
At the same time, the report does not examine all content created or altered using AI. The focus is content that may mislead people about a person, place or event by presenting a materially false or manipulated depiction as real. The report does not generally consider AI-generated text, routine editing or content altered only for aesthetic or technical purposes. The relevant issue in this report is not simply whether AI was used, but whether the resulting content may mislead audiences or cause individual and democratic harms.

This report also does not examine in detail AI-generated child sexual exploitation material (CSEM), which is governed by a distinct legal and policy framework in Australia and raises issues that warrant separate consideration. The Commission recognises that AI-generated CSEM is a serious and deeply concerning issue. Depending on the circumstances, it may involve the exploitation of real children, facilitate ongoing harm to victims, or contribute to the sexualisation and objectification of children more broadly. While these issues are beyond the scope of this report, they raise significant human rights concerns and warrant dedicated attention in their own right.

The two terms, deepfakes and synthetic media, are therefore not interchangeable. All deepfakes considered in this report are forms of synthetic media, but not all synthetic media involves a deepfake. Where the analysis concerns the false depiction of an identifiable person, the report generally uses deepfake. Where it also captures fabricated places, events or other audio-visual content, it uses synthetic media.

1.2 Impact on human rights

Synthetic media is potentially undermining the rights that sustain Australia's democracy. Human rights create the conditions for people to participate in public life, to form views and to hold institutions to account. When synthetic media is used to misrepresent individuals, distort events or manipulate perception, it weakens those conditions in practice.



Synthetic media interferes with privacy by using a person's image or likeness without consent, often in ways that are difficult to detect or reverse. The report shows this most clearly in the growth of deepfake sexual abuse material, which exposes individuals to lasting harm. These practices transform visibility into vulnerability, particularly for women, undermining their ability to engage in public life.

The right to participate in public affairs, protected by art 25 of the *International Covenant on Civil and Political Rights* (ICCPR), depends on people being able to engage without facing disproportionate personal risk. Synthetic media increases the cost of participation by enabling targeted harassment and reputational harm. As these risks fall unevenly across the community, participation becomes more selective, narrowing the range of voices represented in democratic processes.

The impacts of synthetic media are not evenly distributed. The report demonstrates that women, LGBTQIA+ people, culturally and linguistically diverse communities and others are more likely to be targeted and intimidated, humiliated or discredited. Synthetic content often reinforces existing prejudice, amplifying discriminatory narratives and exposing particular groups to heightened risk, which limits their equal participation in public life.

Freedom of expression, protected by art 19 ICCPR, is affected not by direct restriction, but by changes to the conditions in which speech is made and received. Synthetic media may arguably enhance freedom of expression by creating new opportunities for creative communication and participation. When synthetic media falsely attributes words or actions to individuals, however, it weakens confidence in genuine expression and distorts public debate.

Freedom of thought, protected by art 18 of the ICCPR, is challenged by synthetic media introducing forms of influence that operate without transparency. Highly realistic, fabricated content can shape perception without a person's awareness, moving beyond persuasion toward manipulation. Where individuals cannot distinguish between authentic and synthetic material, the ability to form independent views is compromised.

The right to access information, under art 19 of the ICCPR, depends on an environment in which content can be assessed with reasonable confidence. Synthetic media erodes this by blurring the distinction between authentic and fabricated material. As uncertainty increases, people may disengage or become sceptical of all information, including credible sources. This may weaken accountability and the ability to participate on an informed basis.

Decisions about whether to exercise the right to peaceful assembly under art 21 ICCPR are shaped by perceptions of safety and legitimacy. Synthetic media can distort these perceptions by misrepresenting protest activity or those involved, discouraging participation. This constrains collective expression and limits the ability of communities to organise around shared concerns.

Across these rights, synthetic media does not need to be universally believed to have effect. By increasing risk, distorting perception and weakening trust, it reshapes how rights are exercised in practice and, in doing so, places sustained pressure on the democratic system they support.

1.3 Impact on democracy

Australians continue to value democracy deeply, with 73% of Australians reporting in 2026 that democracy is preferable to any other form of government.[7] Yet this support for democracy cannot be taken for granted.[8]

Synthetic media is reshaping democratic life by undermining the human rights that make it possible. Democracy depends on people being able to participate, form views and engage with others in conditions of trust and authenticity. When those conditions are distorted by fabricated content, the effect is not limited to information quality. It alters how rights are exercised in practice, changing who participates, how decisions are made and whether democratic outcomes are regarded as legitimate.

Synthetic media weakens trust in political institutions and those who occupy positions of authority by creating uncertainty about what is real. Deepfakes of public figures can mislead audiences or, equally, enable genuine material to be dismissed as fabricated. As this uncertainty grows, confidence in institutions declines, reducing accountability and making it harder for people to rely on democratic processes and outcomes.

Mapping threats

The United Kingdom's Centre for Emerging Technology and Security has mapped how synthetic content poses a threat to elections across three key stages:

- **Pre-election to create distrust:** Synthetic media can be deployed in early stages of the election period to undermine political candidates or shape voter perspectives on certain topics.
- **Polling period to disrupt processes:** Activities during this period seek to pollute online spaces with volumes of untrue information to confuse people about the election campaign or voting processes.
- **Post-election to discredit results:** Synthetic media may be used to erode trust in the integrity of the election outcome, undermining impact on public confidence in democratic processes more generally.[9]

An independent media acts as the 'fourth estate', meaning it operates alongside the executive, legislature and judiciary to scrutinise power and inform the public. Synthetic media increases the difficulty and risk of this function by complicating verification of information and accelerating the spread of false content. Where inaccurate material is reported or delays occur in verification, public trust in media can decline, weakening its role as a reliable source of information and a check on government power.

Synthetic media provides a low-cost, scalable means of influencing democratic processes. It can be deployed at critical moments to discredit candidates, confuse voters or undermine confidence in electoral outcomes. These activities exploit the speed of online dissemination, reducing democratic resilience by making interference harder to detect and respond to before harm occurs.

Taken together, these impacts show that synthetic media is not a discrete technological issue but a structural challenge to democratic resilience. By weakening trust and enabling covert influence, it alters how democratic systems respond to events. Addressing these risks requires sustained attention to upholding human rights, as these remain the foundation on which democratic legitimacy depends.

1.4 Responding to impacts

The report supports a multi-layered response organised around three reinforcing functions:

- preventing harm where possible
- responding effectively when synthetic media circulates, and
- providing remedies and accountability after harm occurs.

Legal, technical and civic measures each contribute tools and institutional capacity across these functions. The categories are not rigid or strictly chronological. Some measures operate across more than one function, and their effectiveness depends on how they reinforce and are reinforced by the other measures.

Electoral laws can support confidence in voting processes and authorised political communication, provided their role remains targeted and politically neutral. Criminal and online safety frameworks provide accountability, removal pathways and protection against deepfake sexual abuse, and should operate alongside measures that prevent creation, amplification and circulation of deepfakes.

Technical transparency measures, including visible labels and embedded metadata, should be advanced to help audiences, platforms and regulators identify synthetic content. Common standards and careful implementation are needed because effectiveness varies and labels may be removed or applied inconsistently.

Education that supports civic understanding and digital literacy affects whether people can navigate online spaces with greater awareness and make more informed assessments of the content they encounter.

When people are unsure whether content is authentic, trusted public institutions and independent media can provide reliable sources of information and help build confidence in what is true. Efforts to promote accurate information should also include engagement with influencers and other online communicators, reflecting how political information is encountered today.

Synthetic media can undermine democratic participation even if it is not widely believed. It can expose individuals to harm and make people less certain about which information they can trust. These effects may discourage participation and influence whose voices are heard in public life.

Taken together, our analysis supports a multi-layered approach combining targeted legal protections, technical transparency, education, independent media and trusted institutions. Each measure should be directed to the risk it is best equipped to address, while protecting democratic rights and reinforcing the other parts of the response.



1.5 Growing scale and sophistication

AI products are increasingly used to generate synthetic media depicting people or events that did not occur. Advances in generative AI have reduced the technical skill required to produce such material and have increased its availability.[10] In political spheres, 2026 has seen synthetic media being used by Governments across the globe.

Advances in generative AI have not only increased the volume of synthetic media but also changed how it circulates in digital environments. As production becomes easier and distribution more rapid, the impact of synthetic media is also shaped by the systems that determine what people see and engage with online.

Echo chambers

Social media platforms use 'recommender systems' to determine what content people see.[11] Because of these recommender systems, people can be pushed towards influencers and online communities that reflect their own views. Research shows that people are more likely to engage with content that confirms their existing beliefs,[12] which may increase the likelihood that misleading or false content consistent with those beliefs will be accepted or shared.[13]

This can contribute to the formation of 'echo chambers' - digital spaces where people mostly encounter opinions they already agree with, and where disagreement or critical discussion is rare.[14] This can also separate people into different information environments,

eroding the shared facts and common starting points needed for public debate. Democracy depends not only on disagreement and debate, but also on some common understanding of the information being debated. Researchers have argued that synthetic media can have significant consequences for democratic processes by undermining trust in information, distorting political discourse and facilitating manipulation.[15]

Within these echo chambers, synthetic misinformation and disinformation can reinforce existing beliefs and contribute to the spread of conspiracies aimed at eroding trust in institutions, electoral processes, and the integrity of information.[16] This synthetic media undermines the factual foundation on which public debate depends, making it harder for people to agree on what is true.[17] When democratic discourse is no longer grounded in shared facts, it becomes difficult to have meaningful conversations or make informed decisions.[18] Recommender systems that prioritise engagement can then push content based on virality, allowing false narratives to take root – reinforcing bias or inciting hate.[19]

Taken together, these developments show that the risks posed by synthetic media are intensified by the systems through which content is distributed. As engagement-driven platforms concentrate attention around reinforcing narratives, the effects of synthetic media become more persistent and harder to contest. This has serious ramifications for democracy and human rights.

1.6 Building democratic resilience

Australia's existing laws provide an important foundation for responding to synthetic media, but their scope, timing and accessibility vary. Electoral, criminal and online safety laws can deter conduct, support removal and provide accountability, while defamation, copyright, privacy and vilification frameworks may offer redress in particular cases. These pathways are generally reactive and individually focused and should operate alongside preventive and systemic reforms.

Technical transparency measures, including visible labels, embedded metadata and other authenticity measures, should also be advanced. Because detection remains imperfect and labels may be removed or applied inconsistently, common standards and careful implementation are necessary. These measures should support human judgment and rights-respecting legal safeguards, rather than being treated as conclusive proof of authenticity.

Responsibility must be shared across governments, technology companies, media organisations, educators and civil society. The response should operate as a coordinated system in which each measure is directed to the problem it is best equipped to address and reinforces the other parts of the framework.

The risks are already present and will become harder to address as synthetic media becomes more sophisticated and embedded in everyday communication. Australia should not wait for a serious failure of public trust or an election affected by deceptive content before acting.

Australian democracy is currently unprepared for the risks posed by synthetic media. Addressing this challenge will require sustained, coordinated and rights-respecting reform across legal, technical and civic domains.

Findings

Synthetic media is a growing crisis that is interfering with human rights and democratic participation.

It can affect privacy, equality, freedom of expression, freedom of thought, access to information, peaceful assembly and participation in public affairs. The harms associated with synthetic media are often distinctly and disproportionately experienced by women and marginalised communities.

Synthetic media is weakening democratic participation even when the generated media is not widely believed.

By creating uncertainty about authenticity, enabling genuine evidence to be dismissed and undermining confidence in institutions and media, synthetic media makes informed democratic engagement more difficult.

Synthetic media is enabling democratic interference at low cost and significant scale, while making harms difficult to prevent and reverse.

Its speed, reach and potential use at critical moments make manipulation difficult to detect and counter before harm occurs. Even where synthetic content is subsequently corrected or removed, harms to individuals, public trust and democratic participation may endure.

Australia's existing laws provide important but limited protection against the harms of synthetic media.

Electoral and criminal laws may help prevent or deter some harms, while online safety and complaints schemes can support intervention when harmful content circulates. Defamation, copyright, privacy and vilification laws may provide remedies after harm occurs. However, these pathways are generally reactive, individually focused or difficult to enforce at the speed and scale at which synthetic content circulates.

Findings

Measures to prevent, respond to and remedy harm must protect freedom of expression and institutional neutrality.

Responses should address systemic harm without making a government authority the sole arbiter of truth or compromising the neutrality of electoral institutions. Broad prohibitions or takedown powers may capture legitimate satire and commentary, while complaints and removal pathways require clear safeguards and proportionate thresholds.

Technical measures can support transparency, prevention and timely response, but are not a complete solution.

Visible labels, embedded metadata and other authenticity measures should be standardised and interoperable so they can reduce the risk of deception, assist intervention when harmful content circulates and provide information for oversight and accountability. They cannot reliably determine what is real or prevent harmful content from circulating on their own.

Australia requires a coordinated and layered national response that can prevent harm where possible, respond effectively when synthetic media circulates and provide remedies and accountability after harm occurs. No single measure is sufficient.

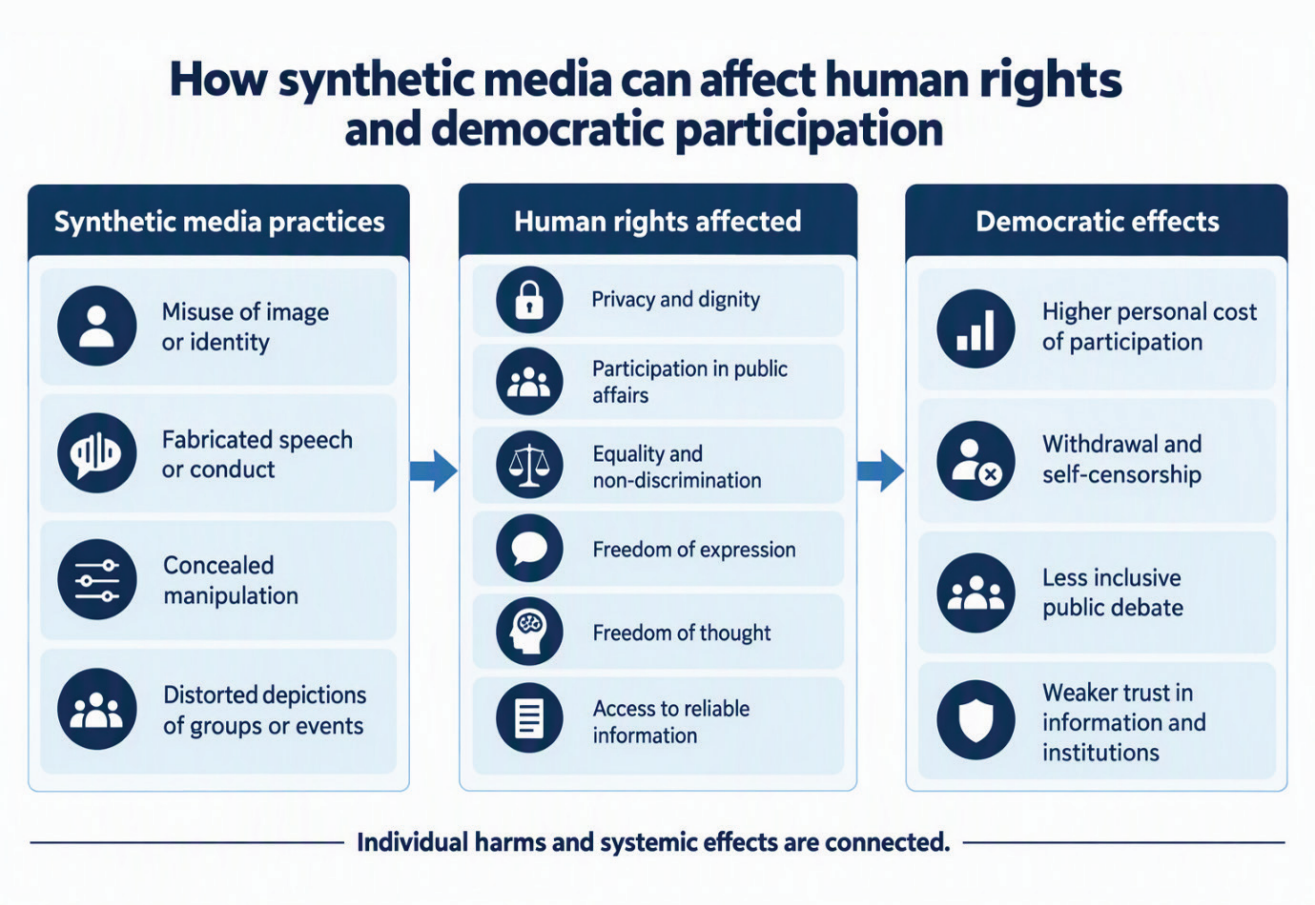
These functions require complementary legal, technical and civic measures, including proportionate laws, technical transparency, digital and civics education, independent media, responsible content creators and trusted public institutions. Measures may contribute to more than one function and should reinforce one another while protecting human rights and democratic participation.

2. Adverse impacts on human rights

This chapter examines how synthetic media can directly affect individuals and interfere with the human rights that support democratic participation. It finds that these harms are not experienced equally, with women and marginalised communities often facing disproportionate targeting and greater barriers to participating in public life. The chapter demonstrates that harms to individuals can also accumulate across society, weakening the shared conditions needed for informed and equal democratic participation.

Finding

Synthetic media is a growing crisis that is interfering with human rights and democratic participation. It can affect privacy, equality, freedom of expression, freedom of thought, access to information, peaceful assembly and participation in public affairs. The harms associated with synthetic media are often distinct and disproportionately experienced by women and marginalised communities.



Synthetic media may affect individuals directly while also weakening the institutions and information environment on which democratic participation depends.

This chapter focuses primarily on harms experienced by individuals. It examines how synthetic media may interfere with a person's rights and ability to participate in public life. Chapter 3 then considers the broader and potentially cumulative effects of synthetic media on democratic resilience. Together, the chapters trace how harms experienced by individuals may, when repeated or amplified across society, place pressure on the conditions that sustain Australia's democracy.

Human rights and democracy are mutually reinforcing, and the harms examined in Chapters 2 and 3 cannot be separated completely. The distinction between the chapters is about focus, rather than a strict division between human rights and democracy.

Australia, like many countries, was one of the many countries that participated in the 2023-2025 global super electoral cycle, where half of the world's population elected local, national and international representatives in 2024 alone.[20] This period of elections came amid a growing shift towards authoritarianism and erosion of human rights protections,[21] as democratic progress stagnated. According to Our World in Data, there are now fewer people living in democratic countries than 15 years ago, with the number of people enjoying democratic rights falling from 3.9 billion in 2016 to 2.2 billion in 2025.[22] While Australia remains a strong democracy, it is not immune from the pressures affecting democratic systems globally, including misinformation, declining trust in institutions and other challenges to democratic resilience.[23]

In the context of this backsliding, it is important to state that human rights and democracy are mutually reinforcing.[24] It is difficult to realise the full suite of rights without democratic institutions and processes, and it is difficult to have a healthy democracy where those rights are not respected and protected.[25]

The Universal Declaration of Human Rights (UDHR) has always been grounded in support for democracy with Article 21 stating:

> Everyone has the right to take part in the government of his country, directly or through freely chosen representatives.
>
> Everyone has the right of equal access to public service in his country.
>
> The will of the people shall be the basis of the authority of government; this will shall be expressed in periodic and genuine elections which shall be by universal and equal suffrage and shall be held by secret vote or by equivalent free voting procedures.

Moreover, the International Covenant on Civil and Political Rights (ICCPR), further reflects this in Article 25:

> Every citizen shall have the right and the opportunity, without any of the distinctions mentioned in article 2 and without unreasonable restrictions:
>
> > To take part in the conduct of public affairs, directly or through freely chosen representatives;
> >
> > To vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors;
> >
> > To have access, on general terms of equality, to public service in his country.

The following rights are especially important and will be closely examined throughout this chapter:

- **Participation in public affairs:** The right to participate in public affairs enables individuals to influence decision-making processes and to choose their representatives freely, ensuring that government authority is grounded in the will of the people.[26]
- **Freedom of expression:** Freedom of expression protects the right to hold opinions and to express political views. This is essential for open debate and accountability in a democratic society.[27]
- **Freedom of thought:** Freedom of thought guarantees individuals the ability to form opinions and beliefs without coercion or interference, which is an important precondition for meaningful democratic participation and independent decision-making.[28]
- **Right to privacy:** The right to privacy protects individuals from arbitrary or unlawful interference with their private life and personal data, which is necessary to ensure free political participation without fear or intimidation.[29]
- **Rights to equality and non-discrimination:** Equality and non-discrimination ensure that all individuals enjoy the same rights and opportunities to participate in public life, preventing the exclusion or marginalisation of particular groups within a democratic society.[30]
- **Access to information:** Access to information enables individuals to understand and scrutinise the actions of public authorities, forming an essential foundation for meaningful political participation.[31]
- **Right to freedom of assembly and association:** Freedom of peaceful assembly and association allows individuals to organise collectively and express their views, which is essential for political participation and the effective functioning of democracy.[32]

Taken together, these rights (and more) demonstrate that democracy depends on a protected civic space grounded in human rights. While every generation of Australians has encountered and overcome challenges to democracy, the rise of synthetic media during a period of democratic backsliding poses serious risk, with the Strengthening Democracy Taskforce noting that there are 'signs of vulnerability' in Australia's liberal democracy.[33] These challenges are not unique to Australia, with democracies globally also grappling with a range of factors including the effects of synthetic media.[34]

2.1 Women's right to participation and privacy

Democratic participation depends on people being able to engage in public life without fear of reprisals.[35] This principle is reflected in the rights to freedom of expression and participation in public affairs protected by arts 19 and 25 of the ICCPR.[36] Public participation is the lifeblood of a thriving democracy, as it supports voting, standing for office, engaging in political debate and joining civil movements. While online platforms have expanded opportunities for civic engagement, they have also introduced new risks. The misuse of synthetic media may facilitate harassment, reputational harm and other forms of abuse that can have a chilling effect on participation in public life.[37]

The rights discussed in this section are universal human rights that apply to everyone. However, the impacts of synthetic media are not experienced equally, and women may face particular and disproportionate barriers to exercising these rights and participating in public life.

Women in positions of public visibility are often disproportionately targeted by gendered disinformation and other forms of online abuse, creating barriers to their safe and effective participation in public life.[38] Attacks involving the deceptive use of synthetic media can

take the form of fabricated videos that portray individuals in compromising or demeaning ways, and are often accompanied by harassment.[39] This behaviour does not just mislead the public, but increases the cost of participation, so that people may be deterred from speaking out or representing their communities in digital public forums.[40]

Targeting journalists

In 2018 journalist Rana Ayyub reported critically about the ruling Bharatiya Janata Party in India. She was subsequently targeted by death threats, deepfake sexual abuse material and doxxing. Ayyub reported experienced anxiety attacks and heart palpitations, and she has since reported censoring herself in light of these attacks.[41]

The case illustrates how synthetic media can be used as part of a broader campaign of intimidation against people who participate in public life. The harm was not confined to Ayyub's privacy or reputation. By contributing to her decision to limit what she said publicly, the campaign also interfered with her freedom of expression and her ability to perform her role as a journalist. Those who find themselves the subject of these kinds of abusive posts often pay a heavy price.[42]

A study by the National Democratic Institute for International Affairs examining politically active women on the social media platform X found that experiencing online abuse was associated with a reduced willingness to remain active on the platform.[43] Further evidence demonstrated this dampens political aspirations of young women and girls, with 25% reporting they felt less confident about sharing their views and 20% reported disengaging from politics due to online misinformation and disinformation.[44]

Harm to careers

Research is starting to reveal the intention of abusers can sometimes be to harm the careers of others, with one abuser stating in a qualitative research interview:

> It was posted in WhatsApp. I was tempted to use some of those other [pornography] websites to ruin her professional career a bit . . . but I didn't do that. I just simply used the personal WhatsApp groups. And some of these groups had up to 50 people.[45]

Although this account reflects the views of one perpetrator, it illustrates the broader risk identified in the study that deepfake abuse may be used deliberately to damage a person's relationships, reputation or professional life. Particularly for women in public-facing roles, the prospect of this harm may increase the personal cost of visibility and participation.

Privacy and sexual abuse

All people have a right to privacy – protecting them from unwanted intrusion into their personal lives. It is recognised under Article 17 ICCPR and includes the right to protect one's image and identity online. Privacy is especially important for the women being subjected to deepfake sexual abuse which uses their likeness without their knowledge or consent. Deepfake sexual abuse is a form of image-based sexual abuse where sexually explicit content is made using someone's image without their consent.[46] Research suggests that women are disproportionately targeted and may consequently experience barriers to participation in public and political life.[47]

Much of the focus in policy and academic literature concerning the impact of deepfakes has been on deepfake sexual abuse material,[48] as deepfake abuse websites attracting substantial levels of user engagement.[49] Women and girls are disproportionately the targets of deepfake sexual abuse,[50] with eSafety Commissioner, Julie Inman Grant, claiming there has been a 550% year-on-year increase in global material since 2019.[51] In the first nine months of 2023 alone almost 150,000 sexualised deepfakes were uploaded - more than all previous years combined.[52]

Cara Hunter

During the 2022 Northern Ireland Assembly Election, Cara Hunter, a Member of the Northern Ireland Assembly was targeted by deepfake sexual abuse material in the form of a pornographic video that was circulated on social media.[53] Hunter describes the moment she found out about the deepfake was 'like watching a horror movie'.[54] As a politician Hunter found it difficult to know how to respond, should she do a press release or hope that so close to an election it simply goes away.[55]

Hunter's own party advised her to ignore it stating 'We're two and a half weeks from an election. If you do a press release, your name will be right up there with words like 'pornography' – and people will see you through a sexual lens and also go looking for it'.[56]

Hunter is now an advocate for reform and is a key voice in Northern Ireland to change laws to better protect people from deepfake sexual abuse material.[57]

Yet the problem is only getting worse due to advancements in AI that make deepfake sexual abuse easier to create.[58] In 2025, changes were made allowing Grok AI to generate deepfake sexualised material on X.[59] Then in late 2025 OpenAI announced that users of ChatGPT would soon be able to generate erotic content – before quietly and correctly scrapping the idea.[60]

The ease of making deepfakes has also meant that the attitudes of abusers are often disturbingly casual, with one man describing his creation of non-consensual deepfake intimate images as a hobby likening it to 'having x-ray glasses' and that it was 'funny'.[61] He went on to state that:

> ... you cannot stop deepfakes. Everyone should live in fear of deepfakes. Every woman on Earth should live in fear of a camera, live in fear of Facebook.[62]

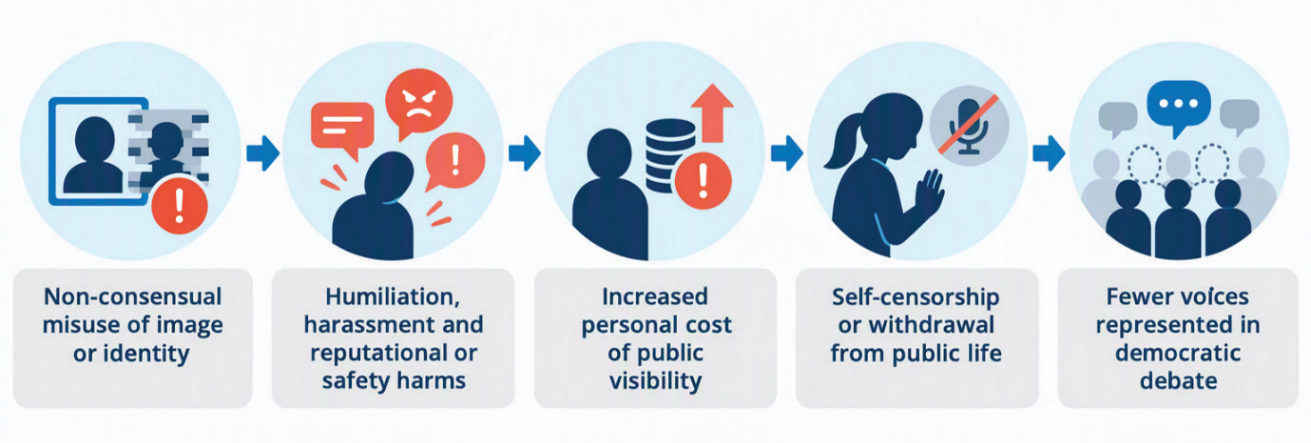
In addition to having rights to privacy and participation - all women also have the right to live free from gender-based violence. This is recognised under the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW). General Recommendation No. 19 first affirmed that gender-based violence is a form of

discrimination, and General Recommendation No. 35 expanded this to include emerging forms of violence, including those facilitated by technology.[63]

Equally, women have a right to be free from inhumane or degrading treatment under Article 7 ICCPR. This right prohibits treatment that causes serious physical or mental suffering or degradation. Deepfake sexual abuse amounts to inhumane or degrading treatment as it can cause psychological and physical harm. The impact is compounded by the permanence and viral spread of online content, which makes removal difficult and re-traumatisation likely.

Finding

Synthetic media is a growing crisis that is interfering with human rights and democratic participation. It can affect privacy, equality, freedom of expression, freedom of thought, access to information, peaceful assembly and participation in public affairs. The harms associated with synthetic media are often distinct and disproportionately experienced by women and marginalised communities.



Taken together, these patterns show that synthetic media is raising the personal cost of participation in ways that fall disproportionately on women. This narrows participation and weakens the representative character of democratic decision-making.

2.2 Equality and non-discrimination

Equal participation in democratic life is not only a democratic ideal but a legal requirement under Article 25 ICCPR. The rights to equality and non-discrimination require that people are able to engage in political activity without being targeted because of protected attributes (e.g. race, disability etc).[64]

Synthetic media threatens this foundation because it intensifies existing patterns of discrimination, frequently drawing upon racialised or discriminatory narratives. Europol (the European Union Agency for Law Enforcement Cooperation) has already warned that synthetic media will be used to incite violence against minority groups, particularly where content is deployed to dehumanise and inflame existing social tensions. [65]

Australian Sikhs targeted

In July 2025 a Facebook post by the page 'Pro Australia' used a synthetic image depicting four turbaned Sikh men holding an Australian Indigenous flag. The accompanying text suggested that some Indian and Sri Lankan Australians were seeking access to benefits intended for Aboriginal and Torres Strait Islander peoples, using the phrase 'our free money'. The post depicted racial stereotypes, portraying these communities as opportunistic and entitled.



Figure: Synthetic image used in a Facebook post by 'Pro Australia', July 2025. Reproduced for the purpose of criticism and review. Source: 'Pro Australia' Facebook post, as reproduced by Amit Sarwal, 'Free money please?': Controversial Facebook post allegedly targeting Indian Australians sparks fury online', The Australia Today (Online, 21 July 2025).

The post attracted more than 300 comments, including hostile and derogatory statements about Indian Australians, although other users challenged the post and highlighted the contribution of Sikh communities to Australia. When synthetic images are combined with misleading claims to target particular communities, it can reinforce existing prejudice and provoke hostility towards those depicted.

As found by the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, synthetic media is being strategically deployed to scapegoat marginalised populations - manipulating public opinion in ways that benefit actors with exclusionary or regressive agendas.[66]

Violence and intimidation

Synthetic media is also being used to promote narratives that normalise online and offline violence against marginalised communities.[67] As noted by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression – when synthetic content reinforces harmful stereotypes, it can embolden individuals to justify violence.[68] History has shown that social media platforms can be weaponised to '… whip up hatred among the population, dehumanise the other side and incite gross violations of human rights'.[69]

The right to life protects every person's right to live and to be free from threats to their life.[70] It is recognised in Article 6 ICCPR, and is non-derogable. This means it cannot be suspended or limited, even in times of war or national emergency. This right forms the foundation for all other rights, because without life, no other rights can be enjoyed.[71] This duty to protect life obliges States to protect people in vulnerable situations because of specific threats or pre-existing patterns of violence.[72]

As found by the Australian Parliamentary Select Committee on Social Media and Online Safety these risks are particularly acute for groups including:

- culturally and linguistically diverse communities
- children and young people
- Aboriginal and Torres Strait Islander peoples.[73]

Equally the right to security of person protects individuals from physical harm and intimidation.[74] It is recognised under Article 9 ICCPR and applies to everyone – regardless of background or identity. These risks are heightened for people from marginalised communities, who already face heightened levels of abuse and discrimination.[75]

The risk of synthetic media being used to incite violence against specific groups is acute during electoral periods, when tensions are heightened and public discourse is more polarised. This kind of polarisation can inflame divisions – inciting violence towards political opponents.[76] In these contexts, hate speech and disinformation can act as catalysts for electoral-related conflict.[77]

Where synthetic media promotes racial hatred or is used to incite discrimination, hostility or violence against a racial, ethnic or religious group, it may also engage Australia's obligations under art 4 of the *International Convention on the Elimination of All Forms of Racial Discrimination* (ICERD) and art 20(2) of the ICCPR. Article 4 requires States parties to condemn and take measures against racist propaganda and incitement, including the dissemination of ideas based on racial superiority or hatred and incitement to racial discrimination or violence. Article 20(2) requires that advocacy of national, racial or religious hatred constituting incitement to discrimination, hostility or violence be prohibited by law.

In some cases, the impacts of deepfakes may be serious enough to engage prohibitions under Australian law, including criminal offences concerning the advocacy or threat of violence against a protected group.

India's 2020 elections

In February 2020, elections were held for the Delhi Legislative Assembly against a backdrop of political and religious tension. During the campaign, AI-generated videos were used to depict political figures delivering campaign messages in languages they had not spoken in the original recordings.[78] Weeks after the election was held, violence broke out during riots in which groups of Hindu's attacked members of the nation's Muslim minority.[79] In the lead up, and during the riots, synthetic content was posted online fabricating scenes of violence and false statements attributed to religious and political leaders.[80] Of the 53 people killed in the riots, two thirds of the victims were Muslims.[81] While the available evidence does not establish that synthetic media itself directly incited or caused the violence, the example highlights the potential for synthetic content to amplify confusion, prejudice and mistrust at a time when reliable information is particularly important.

Taken together, these patterns highlight the risk that synthetic media could potentially discourage particular groups from participating in public life or engaging as fully in public debate.[82] As individuals withdraw from public advocacy or leadership because of this cost, democratic discourse becomes less inclusive and less representative, as the voices most affected by public policy are disproportionately silenced from engaging.[83]

2.3 Freedom of expression and manipulation

Freedom of expression is protected under Article 19 ICCPR and is central to democratic participation because it enables people to contribute to public debate and to challenge the exercise of power. It is the 'foundation stone for every free and democratic society'.[84] Its significance lies not only in protecting individual speech, but in sustaining an information environment in which people can engage with public issues in a meaningful way.

Synthetic media places pressure on freedom of expression by altering the conditions in which expression is produced and understood. By falsely attributing speech or conduct to identifiable people, deepfakes weaken confidence in genuine expression. Equally, synthetic media poses a safety risk to those who are subject to online abuse - casting a shadow over freedom of expression.[85] The fear of being targeted can deter individuals from speaking out or lead them to censor themselves.[86]

Satire as a form of expression

Determining where to draw the line between legitimate expression and unacceptable manipulation is a complex challenge. While some synthetic media can serve as legitimate forms of free expression and commentary,[87] it can also mislead people. Satirical political commentary has a legitimate role to play in democratic expression and should be protected as a form of free expression.[88] Though it can shape the way people think, it is generally perceived as provocative rather than literal.

India's cockroach party

In May 2026, the 'Cockroach Janta Party' began as an online satirical movement representing Indian young people frustrated by unemployment and limited economic opportunities. Its name responded to remarks by India's Chief Justice, Surya Kant, referring to unemployed young people and activists as 'cockroaches'.[89] The movement used AI-generated images and videos of cockroaches dressed as politicians, appearing at podiums and participating in demonstrations.[90] It described itself as the 'Voice of the Lazy and Unemployed', reclaiming the insult as a form of political commentary.[91]



Figure: AI-generated promotional image used by the Cockroach Janta Party. Reproduced for the purpose of criticism and review. Source: Cockroach Janta Party, as reproduced in Zoya Mateen, 'India has a new political superstar—a cockroach', BBC News (online, 21 May 2026).

While the image is clearly unreal, the hyperbolic image of a cockroach politican in a suit while giving a speech for the party demonstrates it's satirical nature as a form of political engagement.

The movement soon extended beyond social media. From June 2026, it helped organise protests over alleged irregularities in India's education and examination systems, with thousands of predominantly young protesters gathering in New Delhi and other cities. By July, the movement had become a significant real-world protest campaign, including a march towards India's Parliament.[92] The movement now has a Cockroach Janta Party website setting out its demands and advocacy.[93]

This synthetic image (along with similar content) can support, rather than undermine, democratic participation. The image is openly satirical and is not intended to deceive audiences about real people or events. Instead, it is about providing young people with an accessible visual language through which to express dissatisfaction, build a shared political identity and mobilise around public issues.

Yet finding the line between a genuine comedic parody (as a form of self-expression),[94] versus genuine synthetic disinformation cloaked under the excuse that it 'was just a joke' is not always so easy.[95]

Anthony Albanese and the CGT memes

Following changes to capital gains tax in the 2026 federal budget, Australian startup founders and small business owners used deepfakes of Prime Minister Anthony Albanese to protest the reforms. The images depicted the Prime Minister working alongside business owners as a co-founder, employee or partner, often claiming that he held a 47% share in their business. This referred to the top marginal income tax rate that some owners argued could apply to gains from the future sale of a business.[96]

The images were generally accompanied by captions making clear that Prime Minister, Anthony Albanese had not actually worked for or invested in the businesses. Prime Minister, Anthony Albanese, responded by thanking their creators for choosing 'very nice photos' of him, before defending the proposed reforms.[97]



Figure: AI-generated image depicting Prime Minister Anthony Albanese as a plumber. Reproduced for the purpose of criticism and review. Source: as reproduced in Tess Bennett, 'Albanese mocked by entrepreneurs using AI images to fight CGT changes', Australian Financial Review (online, 18 May 2026).

The deepfake is an accessible way in which people are engaging in political discourse (even if it does oversimplify policy discussions) while also highlighting the challenges of distinguishing satire from deception. Although the images used the Prime Minister's likeness without depicting real events, their exaggerated context could lead audiences to identify them as political satire rather than factual representations. The case shows that context matters when assessing synthetic media. The same technology can be used for different purposes and there may be a real, if difficult to define, distinction between deceiving audiences and enabling democratic participation.

This distinction between satire and deceptive information is especially relevant in the current political climate, where both major parties in the 2025 Australian federal election have experimented with political memes and rudimentary deepfakes (also known as 'shallowfakes'). This is in addition to similar global trends where deepfakes are being used to negatively depict or mislead people.[98]

There is a risk that such use may soon create confusion amongst voters who place trust in messaging by authorised political parties. While political synthetic media should not be used to mislead voters, there are also genuinely positive uses as a form of political commentary.

2.4 Interference with peaceful assembly

The right to freedom of peaceful assembly under Article 21 ICCPR enables democratic participation beyond formal electoral processes. The right means that people can come together in public or private spaces to protest or support a cause. This right is a cornerstone of any healthy democracy because it allows people to make their voices heard.

In the digital age, the exercise of peaceful assembly depends in part on online spaces to facilitate protests or make decisions about whether to attend.[99] Synthetic media can undermine these conditions. False synthetic imagery about protests can be rapidly disseminated to distort perceptions of protest activity to discourage engagement, discredit movements and promote stronger police responses. Synthetic media may be used to prevent people from participating by spreading fake information - from where the protest is happening, how dangerous it is or what the protest is about.[100] Alternatively deepfakes can also target protest movements by fabricating videos that depict organisers or participants engaging in illegal or violent behaviour.

Synthetic media is also being used to provoke or intensify sometimes violent protest responses to certain events. For example, following the murder of Henry Nowak and an unrelated later stabbing attack in Belfast, the United Kingdom faced mass protests and demonstrations.[101] These protests may have been partly influenced or exacerbated by synthetic media being posted online to stoke anti-migrant sentiment - depicting migrant men assaulting white women or replacing iconic British buildings with large mosques.[102] The Southport riots in 2024 in the UK were also influenced by synthetic media according to analysis by The London School of Economics and Political Sciences.[103]

Closer to home, synthetic media depicting a massive crowd at the 2025 March for Australia had to be addressed and fact checked by the Australian Associated Press (AAP).[104] In a separate example, an AI-generated image was shared on Facebook with the false claim that it showed a police officer using force against a protester at a pro-Palestine rally in Melbourne. Although a genuine confrontation between police and protesters had occurred, the depicted scene did not appear in footage of the event and the officer's uniform did not match that of Victoria Police.[105]

By distorting information about protest activity, synthetic media may influence people's decisions about whether and how to participate in protests. It may also shape how the broader public perceives particular protests and the causes associated with them. Where participants rely on distorted information about safety, popularity or legitimacy, the practical ability to exercise this right is reduced. This weakens collective expression and limits a key pathway through which people participate in democratic life.

2.5 Manipulation of free thought

Under Article 18 ICCPR, freedom of thought is the right to hold your own beliefs, ideas and opinions without interference.[106] In a democracy, freedom of thought allows people to think independently, question what they're told and explore different viewpoints. Despite its importance the right is underdeveloped and poorly understood,[107] often described as the 'forgotten freedom'.[108]

In the digital age, the right to freedom of thought is increasingly vulnerable to subtle and sophisticated forms of manipulation. While much political content online should be viewed as mere persuasion, synthetic media can move away from this and into an impermissible alteration of thoughts.[109] Drawing on a framework proposed by the Special Rapporteur on freedom of religion or belief,[110] synthetic political content raises concerns about manipulation where individuals are unaware of attempts to influence them, where such attempts are concealed, or where significant information or power asymmetries exist.

The three elements set out in the framework can help to explain when influence through synthetic media may move beyond ordinary political persuasion towards manipulation that interferes with the right to freedom of thought:

- **Lack of consent:** Although users may technically consent to use the social media platforms that often house synthetic media, they are not providing specific informed consent to being shown such content or attempts to manipulate them through synthetic media.
- **Concealment:** Although it would vary on a case-by-case basis, where synthetic media is presented as authentic and not identifiable as AI generated, attempts to influence audiences may be concealed from those viewing the content. This would lead a reasonable person to not be aware of the intended attempt to influence.
- **Asymmetrical power:** It may be difficult to assess the power dynamic between the influencer (someone posting synthetic political content) and rights holder, yet it is possible that there is some power imbalance due to the perceived authority of posts containing hyper-realistic deepfakes (which can be presented as news).[111]

Synthetic media moves beyond influencing opinion to interfering with the conditions under which thought is formed. Where content is designed to deceive without awareness or consent, it disrupts the ability to form independent views. This undermines a foundational democratic safeguard by weakening the integrity of individual judgment on which collective decision-making depends.

2.6 Access to reliable information

The right to access information is protected under Article 19 ICCPR, which guarantees the freedom to seek, receive and impart information. This right underpins democratic engagement by enabling people to understand public issues and undertake informed debate. In digital environments, its effective exercise increasingly depends not only on the availability of information, but on whether that information can be assessed as reliable in practice.

Digital technologies have fundamentally changed how information circulates in public life, reshaping how people learn about events and engage in democratic processes.[112] These developments have expanded access to information and increased opportunities for participation, while also raising expectations about the reliability of information online.[113] In a society where political debate and civic engagement increasingly occur online, the integrity of information has become a practical condition for the exercise of democratic rights.

Information integrity refers to an ecosystem in which reliable and accurate information is available to everyone - enabling people to engage meaningfully in public life and exercise their rights.[114] Strengthening information integrity means protecting the right to freedom of expression and access to information and enabling people to navigate information spaces safely.[115]

Information integrity does not require uniformity of views or the absence of error. Rather, it reflects whether the online environment enables people to evaluate information without their understanding being undermined by fabricated or misleading content. Where information integrity is eroded, rights may remain formally protected but become progressively harder to exercise in practice.[116]

Recent technological developments have placed this condition under increasing strain.[117] While digital platforms allow information to be disseminated rapidly and widely, they have also enabled false or misleading content to circulate at a scale that challenges established methods of verification.[118] Advances in AI have further weakened traditional signals of authenticity.[119] Synthetic media can now be presented in ways that mimic evidence, introducing uncertainty not only about what is false, but about whether any information can be relied upon.

Against this backdrop the United Nations Global Principles for Information Integrity provides a rights-based lens for improving the information ecosystem and protecting the right to information. The five principles are:

- **Societal trust and resilience:** recognises that information integrity relies on public confidence in the credibility of information and on society's ability to absorb and respond to manipulation without democratic participation being weakened.[120]
- **Healthy incentives:** examines how the design and monetisation of platforms shape information flows, particularly where commercial incentives encourage the amplification of viral synthetic media.[121]
- **Public empowerment:** focuses on whether people are able to engage with information environments in ways that support informed judgment.[122]
- **Independent, free and pluralistic Media:** affirms that information integrity depends on the presence of media institutions that operate independently of undue influence and provide reliable information.[123]
- **Transparency and research:** addresses the need for visibility into how information systems function and for protected research that can identify emerging risks.[124]



As uncertainty about what is real increases, the practical value of access to information is diminished. Where people cannot assess the credibility of what they encounter, their ability to participate in public life on an informed basis is weakened. This erosion of information integrity therefore reduces the effectiveness of democratic engagement and limits the capacity to hold institutions to account.

3. Democratic resilience

This chapter examines the broader and cumulative effects of synthetic media on democratic resilience. It finds that synthetic media can distort public debate and weaken shared confidence in information and institutions - even when the content is not widely believed. The chapter concludes that protecting democratic resilience requires a coordinated, multi-layered response.

Finding

Synthetic media is weakening democratic participation even when the generated media is not widely believed. By creating uncertainty about authenticity, enabling genuine evidence to be dismissed and undermining confidence in institutions and media, synthetic media makes informed democratic engagement more difficult.

Finding

Synthetic media is enabling democratic interference at low cost and significant scale, while making harms difficult to prevent and reverse. Its speed, reach and potential use at critical moments make manipulation difficult to detect and counter before harm occurs. Even where synthetic content is subsequently corrected or removed, harms to individuals, public trust and democratic participation may endure.

Democratic resilience refers to the capacity of democratic systems to withstand disruption and continue to function effectively under changing conditions.[125] It depends on people being able to participate in public life in ways that support accountability, legitimacy and public confidence in democratic institutions. In this sense, democratic resilience is grounded in the effective enjoyment of human rights, including freedom of expression, access to information and participation in public affairs (amongst others).

Synthetic media affects the conditions that sustain democratic resilience in practice. It may weaken trust in public institutions and political figures, complicate the capacity of the media to perform its scrutiny and verification function, and increase the role of other entities in shaping information and pubic debate. Synthetic media does not need to be widely believed to have impact. It can be effective simply by creating uncertainty about what information can be trusted reducing people's ability to remain confident and engaged in democratic processes.

3.1 Erosion of trust in institutions and figures

Historically, Australia enjoyed some of the highest levels of political trust globally.[126] However, over the past decade, this trust has steadily declined - mirroring a broader international trend of democratic backsliding.[127] Political trust reached record lows in 2019 with only 25% of voters trusting government, yet the latest evidence shows a small improvement to 30%.[128] There is a risk that, when distrust reaches a certain point, people may disengage from public debate or become less willing to rely on information from otherwise credible sources.

Synthetic media is blurring the line between truth and fabrication. When people can no longer trust what they see or hear, it becomes nearly impossible to build a shared understanding of facts.[129] The Select Committee on Adopting Artificial Intelligence has warned that synthetic media has the potential to erode trust in public institutions,[130] which can undermine confidence in democracy itself.[131]

India's 2024 Election

During India's 2024 general election, deepfake videos falsely depicted Bollywood actors Aamir Khan and Ranveer Singh criticising Prime Minister Narendra Modi and encouraging viewers to vote for the opposition Congress party.[132] The videos appeared to show the actors accusing Modi of failing to address economic issues. Both ended with the Congress party's election symbol and the slogan 'Vote for Justice, Vote for Congress'. The actors confirmed that the videos were fake.[133]

One post sharing the Ranveer Singh video received 438,000 views, 8,700 likes and 2,900 reposts.[134] Although the post was labelled as manipulated media by X, the person who shared it did not remove it because she believed the content still had creative value.[135]

The case shows how a public figure's identity can be used without consent to give credibility and reach to a political message. It also demonstrates that labels and corrections may not prevent manipulated content from circulating widely during an election. While there is no direct evidence that these particular videos changed how people voted, their scale and political content illustrate the potential for deepfakes to distort election debate.

Trust in institutions

Governments are also beginning to use synthetic spokespeople in routine public communications. In 2024, Ukraine's Ministry of Foreign Affairs introduced Victoria Shi, an AI-generated spokesperson used to deliver official statements on consular matters, while human officials remained responsible for preparing and verifying the content. Each statement included a QR code linking to an official text version as a safeguard against impersonation and manipulation.[136]

These tools were intended to make government communication more accessible and engaging. However, their use may also have longer-term implications for institutional trust. As synthetic representatives become part of everyday interactions with government, clear disclosure and reliable ways to authenticate official communications will be important to ensure that people know when they are engaging with synthetic content and who remains responsible for what is communicated.

Trust in the administration of justice

The risks posed by synthetic media also extend to the administration of justice and public confidence in police, courts and other legal institutions. Synthetic media can be used to fabricate evidence, impersonate public officials, disseminate false information about police investigations or legal proceedings, and create uncertainty about the authenticity of genuine evidence. As these technologies become more sophisticated and accessible, they may increase the resources required to verify information, undermine confidence in institutional decision-making, and complicate investigations and the administration of justice.

Following the December 2025 Bondi terrorist attack, a deepfake video appeared online depicting Australian Federal Police Commissioner Krissy Barrett announcing that four Indian nationals had been arrested in connection with the attack.[137] One post of the video received more than 130,000 views before the claim was debunked by fact-checkers.[138] The incident demonstrates how synthetic media can exploit public reliance on trusted authorities during periods of heightened community concern, potentially undermining confidence in official information and creating confusion about the status of investigations.

Although the implications of synthetic media for investigations and legal proceedings are beyond the scope of this report, the technology raises broader questions about the administration of justice. These include how justice institutions can maintain public confidence in the authenticity of evidence and official communications, respond effectively to synthetic media that misrepresents institutional actions or decisions, and preserve trust in the fairness and integrity of justice processes in an increasingly synthetic information environment.

Trust in key figures

Public trust in political figures is fundamental to the legitimacy of democratic institutions and processes. When elected officials are seen as accountable and transparent, it strengthens confidence in the broader systems of governance. Conversely, when people believe that politicians can act without consequence, trust begins to unravel. Deepfakes of politicians can create confusion about electoral processes and undermine trust in leaders.[139] They may raise questions about whether people can trust their own eyes and ears when watching political figures speak.[140]

Charlie Kirk Memorial Video

Following the assassination of US political commentator Charlie Kirk, deepfakes circulated widely online. At a church service in Texas, an AI-generated audio clip depicted Kirk commenting on his own death and urging listeners to continue advancing the causes he had supported.[141] Other synthetic images depicted Kirk with religious symbols or figures, contributing to a broader portrayal of him as a martyr.

Synthetic media was also used to mock Kirk. In a trend known as ‘Kirkification’, his face was inserted into unrelated memes, reaction images and videos. Most of this content was not intended to deceive. However, the contrasting uses of Kirk’s likeness demonstrate how a person (and their estate) can lose control over how their identity and legacy are represented after death.

The example also illustrates how deepfakes can shape political memory. It may help supporters express grief and strengthen a shared political identity, but it can also simplify a contested public figure into either a martyr or an object of ridicule. Where synthetic content is difficult to distinguish from authentic material, it may narrow the space for a more critical and nuanced discussion of that person’s life, views and legacy.

Declining trust in political figures is deeply concerning, yet increasingly predictable in an era where deepfakes are being used by politicians themselves.[142] For example, US President Donald Trump has repeatedly used deepfakes and synthetic media to communicate his political message to followers online. His posts have displayed everything from him fighting the Canadian Men's Ice Hockey Team to a post depicting him as Jesus-like figure.[143]

Obama arrested for stealing election

In July 2025 US President, Donald Trump, posted a deepfake on Truth Social showing former President Barack Obama being arrested in the Oval Office and subsequently imprisoned for allegedly stealing the 2016 election.[144]



Figure: AI-generated image of Former US President Obama from President Trump's Truth Social Post. Reproduced for the purpose of criticism and review. Source: President Trump's Truth Social account, as reproduced in Joe Sommerlad, 'Trump posts AI video of Obama being arrested after Gabbard's coup claims: 'No one is above the law'', Independent (Online, 22 July 2025).

The post followed claims by Director of National Intelligence Tulsi Gabbard that members of the Obama administration had been orchestrating a 'years-long coup'.[145] The video placed those contested allegations within an emotive visual narrative in which a former president was portrayed as a criminal being punished.

This deepfake is being used to reinforce an existing political narrative by presenting an imagined outcome as if it had occurred. Even where audiences recognise the content as fabricated, its circulation by a serving president may legitimise calls for action against a political opponent and deepen hostility between political groups. It also illustrates how synthetic media can reduce a complex and contested issue to a simple visual depiction of guilt and punishment.

When high-profile political figures create or endorse deepfake content, it forces the public into a constant state of scepticism. While digital literacy is essential, the sheer volume of manipulated content makes it increasingly difficult for individuals to assess what is real.

Deepfake-enabled scams are another area of significant concern. In Australia, scammers have used the images and voices of well-known public figures, including mining magnate Andrew 'Twiggy' Forrest and financial journalist David Koch, to create false endorsements of cryptocurrency and investment schemes.[146] The apparent involvement of trusted public figures can make these scams more convincing, causing financial harm to individuals while also undermining trust in legitimate public communications. Scamwatch has warned that fake news articles and deepfake videos promoting online investment platforms are increasing on social media.[147]

The use of deepfakes in financial scams falls outside the principal human rights and democratic participation focus of this report. However, these examples demonstrate the broader harms that may arise from the unauthorised use of a person's image - and the growing difficulty of distinguishing genuine endorsements from fabrications. Further work is needed to assess whether Australia's consumer protection, financial regulation and platform-governance frameworks can respond effectively to the increasing scale and sophistication of deepfake-enabled scams.

Liar's dividend

An environment of mistrust enables political actors to exploit the so-called 'liars dividend' - the ability to dismiss genuine evidence as fake.[148] As deepfakes proliferate, politicians can claim that real footage is synthetic, further eroding public confidence.

The Murder of George Floyd

In June 2020, Winnie Heartstrong (a Republican candidate for the US House of Representatives) published a report falsely claiming that footage of George Floyd's murder was a deepfake that had been staged to inflame racial tensions.[149] The posting is no longer available on the website 'investigatedeepfakefloyd.com'.

Heartstrong presented the claim as genuine and called for a federal investigation.[150] The allegation contradicted eyewitness recordings, surveillance footage and official autopsy findings confirming Floyd's death.

The example illustrates the 'liar's dividend' as growing awareness of deepfakes can be used to cast doubt on genuine evidence. No synthetic media was needed. Merely claiming that authentic footage was fabricated allowed a public figure to promote a false account of a documented event and potentially weaken confidence in visual evidence.

The liar's dividend may also have the effect of shifting the burden of proving content is real onto journalists and human rights advocates.[151] This increases pressure on media and human rights organisations to counter claims that evidence of human rights abuses aren't real.[152]

Benjamin Netanyahu 'proof-of-life' videos

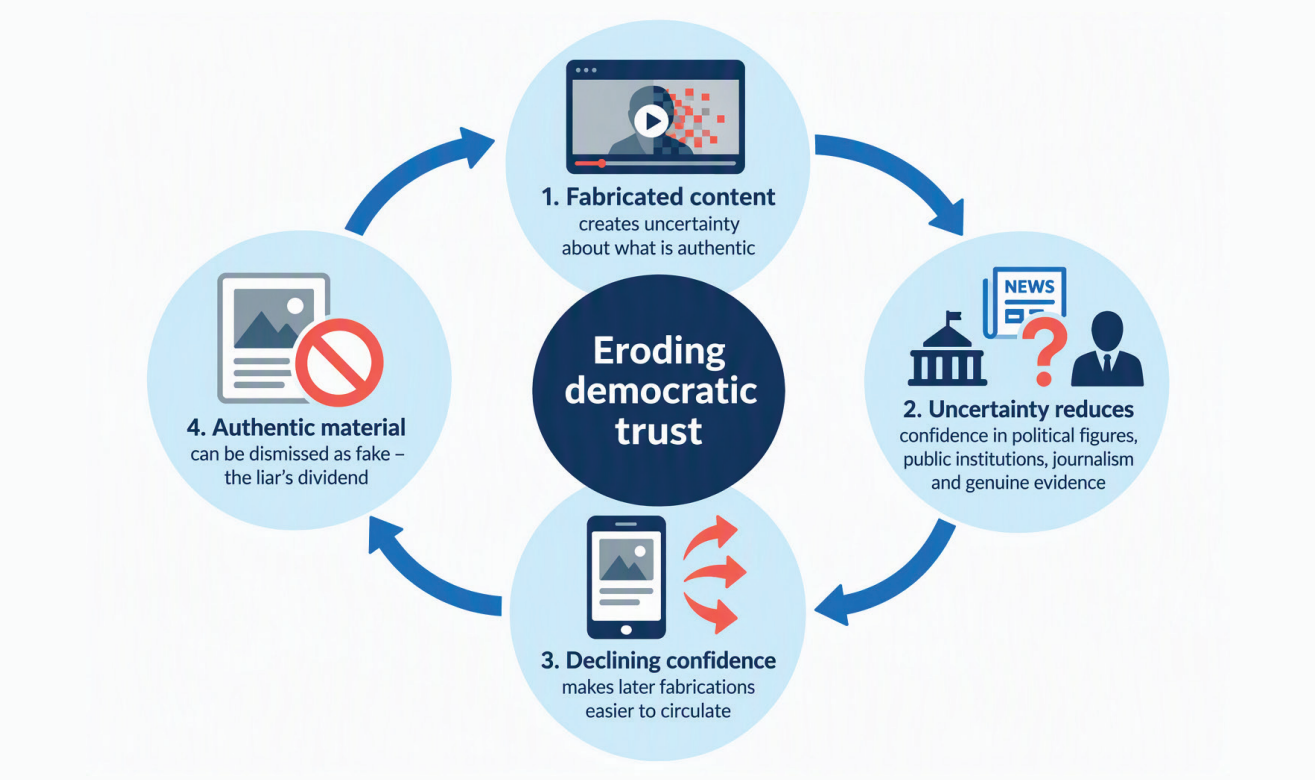
During the 2026 conflict with Iran, false claims circulated online that Israeli Prime Minister Benjamin Netanyahu had been killed. In response, the Israeli Prime Minister released videos showing that he was alive. However, these videos were themselves alleged to be AI-generated, with social media users pointing to supposed visual anomalies as evidence. Fact-checkers found no credible evidence that the footage was synthetic or that Netanyahu had been killed.[153]

The example illustrates the liar's dividend because the growing awareness of synthetic media made it easier to dismiss apparently genuine evidence as fabricated. Rather than resolving the false claims, the release of additional footage generated further speculation about its authenticity. The result is not only the spread of false information, but declining trust in images and videos that are real.[154]

These developments show how synthetic media is weakening societal trust in ways that reduce collective resilience. When fabricated content circulates alongside genuine material, uncertainty spreads beyond individual claims and begins to affect confidence in political figures and democratic institutions. As this uncertainty accumulates, society becomes less able to handle manipulation without disengagement or cynicism taking hold.

Finding

Synthetic media is weakening democratic participation even when the generated media is not widely believed. By creating uncertainty about authenticity, enabling genuine evidence to be dismissed and undermining confidence in institutions and media, synthetic media makes informed democratic engagement more difficult.



These developments highlight that declining trust is not simply a matter of public perception, but a consequence of weakened conditions for the exercise of fundamental rights. Where the information available to people to understand and evaluate the actions of those in power is unreliable, their ability to seek, receive and impart information consistent with Article 19 ICCPR is diminished. This can reduce confidence in institutions and limit society's capacity to recognise and resist manipulation without becoming disengaged, weakening the human rights foundations on which democratic legitimacy depends.

Media as a fourth estate

An independent media acts as the 'fourth estate', meaning it operates alongside the executive, legislature and judiciary to scrutinise power and inform the public. Media undertakes this role within a widely understood ethical framework, which helps distinguish professional journalism from information shared by individual citizens. Journalism shapes how people understand public events and evaluate the actions of those in positions of authority, which in turn affects the quality of democratic resilience. This function depends on media being able to operate without undue interference and on journalists being able to work safely and with accuracy.[155] Synthetic media complicates verification of information by journalists and audiences and increases the risk that synthetic content is circulated as credible information.

The way Australians consume news has undergone a dramatic shift over the past decade. Until recently most Australians received information about current affairs through TV, radio or newspaper.[156] Legacy media created the shared reality upon which all Australians were able to understand the world around them. However new media (podcasts, social media, YouTube) is rising in prominence creating a more fragmented information environment.[157]

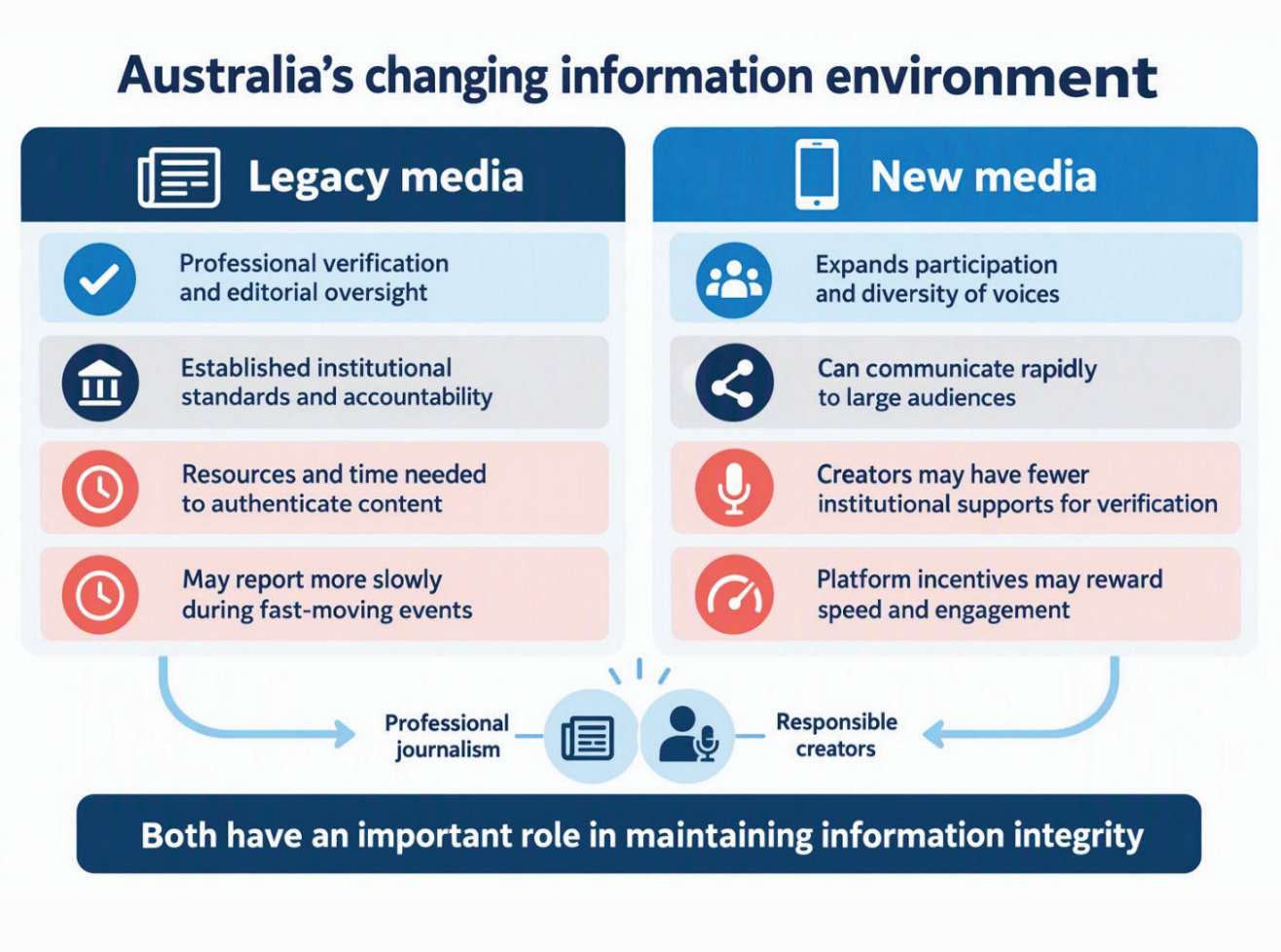
The audiences for these newer forms of media can be substantial. In June 2026, 55% of Australians reportedly consumed podcasts each month, while news podcasts collectively reached almost 4.7 million monthly listeners.[158] These figures illustrate that podcasts and other creator-led formats are no longer peripheral to Australia's information environment.

New media has transformed the information landscape, enabling broader participation in public discourse and facilitating the free flow of ideas – an essential feature of a healthy democracy.[159] However, these same platforms have also become vehicles for the spread of synthetic media.[160] In this fragmented media environment, it has become increasingly difficult for people to discern which sources are trustworthy.[161]

Content is now shared more rapidly and cheaply than ever, often prioritising attention and engagement over accuracy.[162] Unlike legacy media journalists (who typically receive formal training in verifying facts and assessing credibility) many content creators and influencers lack the resources or editorial oversight to ensure the integrity of the information they share.[163] They may also operate within environments where commercial and platform incentives reward speed and volume over nuance and investigation. That isn't to say creators do not ever verify information or check sources – however they do not have the same institutional supports as legacy media does.

While television and legacy media remain the most popular sources of news (37%), social media platforms (26%) have now overtaken online news sites (23%) for Australians.[164] While Facebook is the most commonly used platform for news,[165] among younger Australians aged 18–24, TikTok has emerged as a key source of news, used by more than one-third of this age group.[166]

As short-form video content (usually on TikTok, Instagram, Facebook etc) becomes a primary channel for news consumption, the risk of encountering synthetic content increases. This is especially concerning as 72% of Australians believe most people don't understand when information is misleading or fake during elections.[167]



Legacy media once served as trusted gatekeepers of information, helping to uphold standards of truth and accountability in democratic societies.[168] Synthetic media makes their job more difficult.[169] For example, journalists must now deploy methods to determine if videos and images are synthetic.[170] This can prevent legacy media from rapidly reporting on real and urgent events – or when they do report and accidentally publish synthetic media it can result in further loss of trust.[171]

Fake speeches in congress

In August 2025, American journalist and TV anchor Chris Cuomo posted a deepfake video of Congresswoman Alexandria Ocasio-Cortez arguing on the floor of Congress that a recent Sydney Sweeney ad for American Eagle was racist. Cuomo posted the deepfake with the caption:

> Nothing about hamas or people burning jews cars. But sweeney jeans ad? Deserved time on floor of congress? What happd to this party? Fight for small business ... not for small culture wars.[172]

Ultimately it was established the speech never occurred, and the Congresswoman responded that the video was not real. When deepfakes are shared by new or legacy media, it can undermine public trust in journalism.[173]

While some influencers contribute to the erosion of trust by sharing synthetic media, others are actively working to restore trust by promoting accurate, timely and transparent information.

Influencers challenging synthetic media

Dylan Page is a social media news creator whose short-form videos have attracted a large online audience. In 2026, Page published an investigative video titled 'I Tracked Down the Stranger Who Stole My Identity', describing how an unknown person had used AI to reproduce his face, voice and presentation style. The synthetic presenter appeared in videos published through an account called 'NewsBaby', which replicated the format of Page's news content and reportedly received millions of views.

Page traced and met the person behind the account, documenting how the synthetic version of his identity had been created. His investigation showed how readily a creator's publicly available content can be used to reproduce their appearance, voice and manner of communicating at scale. Page's video had received more than 600,000 views at the time of writing, demonstrating that creator-led investigations can themselves reach substantial audiences.

The example illustrates that new media creators are not only potential sources or distributors of synthetic content. Some are also using their reach and familiarity with online platforms to explain how synthetic media is created, expose deceptive practices and encourage audiences to assess online content more critically.

As noted by the Grattan Institute, 'reliable news and journalism are at the heart of public discussion, diversity of voice, open justice, accountability, and informed decision-making'.[174] A consequence of the growth of new media has been the more rapid spread of false information online.[175]

Trust in news in Australia has fallen to just 43%,[176] reflecting a broader decline in confidence as access to news across all platforms continues its long-term downward trend.[177] Synthetic media increases this distrust, as viewers must become more sceptical of what they see and hear. Consequently, people are becoming more wary of even legitimate information from trustworthy news sources.[178] The proliferation of synthetic media masquerading as news harm the integrity of journalists and influencers – eroding their credibility and weakening the public's confidence in the information they receive.[179]

The right to seek, receive and impart information relies not only on access to information but also on confidence in the source of the information. When deepfakes and other forms of synthetic media undermine trust in media and other information sources, people may become less able to make informed decisions, participate in public debate and hold those in power to account. Over time, this can weaken both the practical enjoyment of the right to information and democratic participation.

3.2 Interference in Australian democracy

Social media plays a complicated role in Australia's democracy. It can be a powerful tool for good - helping people access information and take part in public debates but also comes with risk. For example, the Senate Selection Committee on Foreign Interference through Social Media observed:

> Whether we like it or not, social media platforms are today not just the dominant communication channels in modern economies, they constitute the public square for democracies. They are the place where news is first reported, contentious issues are debated, consensus is formed and public policy decisions are shaped. The health of these forums directly affects the health of our democracies.[180]

Increasingly foreign actors have realised that social media is a cheap and effective way to influence global politics – often weakening our democracy in the process.[181] In recent years, there's been a noticeable rise in interference online - especially during elections and referendums.[182] In its 2020-21 Annual Report, the Australian Security Intelligence Organisation (ASIO) said that espionage and interference was a principal security concern as '[m]ultiple foreign governments are determined to interfere in Australia's democracy and undermine our sovereignty'.[183] ASIO's most recent annual report states that 'extreme levels of espionage and foreign interference did not diminish' in 2024–25, with multiple nations continuing efforts to interfere in Australia's political processes.[184]

ASIO's 2024-2025 Annual Report has also observed that overseas conflicts and tensions are amplified through social media and online echo chambers - contributing to reduced social cohesion and greater intolerance in Australia.[185] These observations demonstrate how online environments can magnify security threats and their effects on Australian communities.

This comes at a time when anti-democratic actors seek to sow political violence in Australia.[186] ASIO Director-General Mike Burgess has warned that:

> Many of the foundations that have underpinned Australia's security, prosperity and democracy are being tested: social cohesion is eroding, trust in institutions is declining, intolerance is growing, even truth itself is being undermined by conspiracy, mis- and disinformation.[187]

Bondi terror attack

In the immediate aftermath of the Bondi antisemitic terrorist attack on 14 December 2025, false and misleading content spread rapidly online, often before verified reporting was available. This included synthetic images and fabricated profiles that misrepresented people connected with the attack.[188]

Arsen Ostrovsky, who was shot in the head and seriously injured in the attack, became the subject of AI-generated images suggesting that he had staged his injuries. ABC News Verify confirmed that the images purporting to show his injuries being staged were AI-generated.[189]

This included synthetic images falsely depicting the attacker in fabricated scenarios, as well as manipulated content used to construct misleading narratives about identity and motives.[190]

Ostrovsky told the Royal Commission on Antisemitism and Social Cohesion that the material began circulating while he was in hospital preparing for surgery. He described the cumulative harassment as 'overwhelming' and 'dehumanising', saying that it sought to erase his experience and trauma.[191]

Further examples included altered images of victims and fabricated profiles that were used to support conspiracy claims or falsely implicate individuals.[192]

The case illustrates two connected harms: the direct harm caused to an individual when their image and experience are manipulated, and the broader harm caused when synthetic media is used to cast doubt on genuine evidence and promote a false account of events.

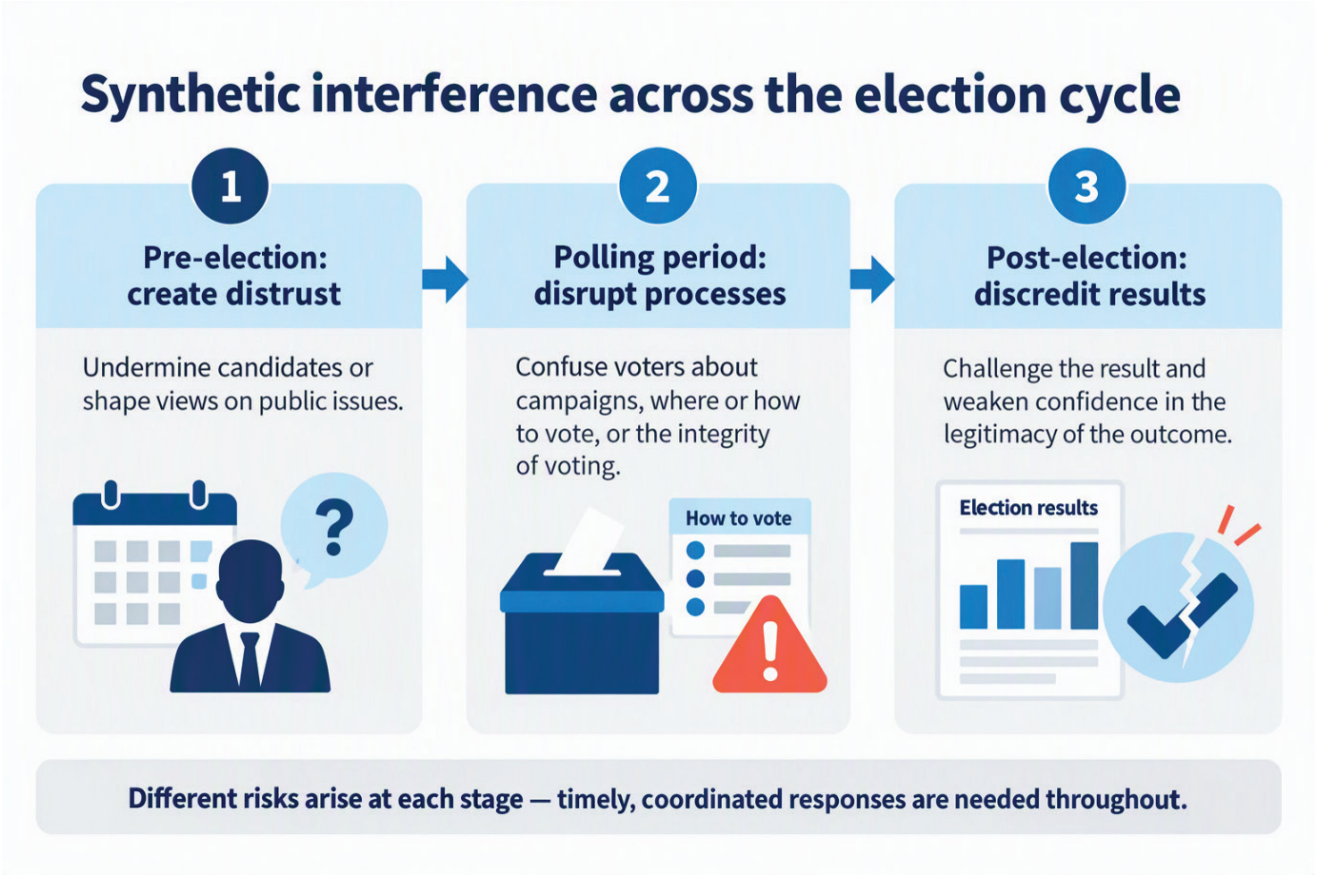
Social media and its recommender systems can magnify these risks by rapidly promoting sensational or inflammatory content to wider audiences.[193] It is by this process that entities can introduce inflammatory material, which is then promoted by algorithms, encouraging further user engagement and amplifying the reach of the content.[194] For foreign entities seeking strategic advantages from creating democratic instability in a rival country, synthetic media presents significant opportunities.

Spread via social media, synthetic media can quickly reach millions of people in a short time allowing malicious actors to:

- attack a candidate's image and credibility
- harass or humiliate those engaged in democratic discourse online
- manipulate public opinion
- incite violence against minority groups
- stoke social unrest and political polarisation.[195]

The Centre for Emerging Technology and Security has mapped how synthetic content poses a threat to elections across three key stages:

- **Pre-election to create distrust:** Synthetic media can be deployed in early stages of the election period to undermine political candidates or shape voter perspectives on certain topics.
- **Polling period to disrupt processes:** Activities during this period seek to pollute online spaces with volumes of untrue information to confuse people about the election campaign or voting processes.
- **Post-election to discredit results:** Synthetic media may be used to erode trust in the integrity of the election outcome, which has a serious impact on public confidence in democratic processes more generally.[196]



2024 US Election

During the 2024 US election, a deepfake of a 'former student' was created alleging that Democratic Nominee for Vice President, Tim Walz, had abused him while teaching.[197] The deepfake was sophisticated because although the video was fake, the person depicted was real and had been a student at Walz's school.[198] It is estimated that the initial post was viewed over 5.4 million times within 22 hours of going live on X.[199]

The deepfake was allegedly attributed to the Russian-aligned network Storm-1516.[200] It was likely intended to discredit and attack Walz to manipulate the views of voters.

Europol (Europe's agency for law enforcement) has stated that synthetic media could be used to impersonate public officials to interfere with elections or encourage unrest.[201] The rapid dissemination of synthetic media can allow people to strategically post content close to an election – taking advantage of the limited time for targets to respond and counter the false claims.[202] This is especially concerning given the increasingly long pre-polling periods in Australia which provide a large window of time for synthetic media to be deployed.[203]

For example, in Australia a malicious actor could release a deepfake within the extended pre-polling period which impacts people's voting decisions – this would be incredibly difficult for a targeted candidate to address given the speed with which deepfakes spread online.[204]

A deepfake of a candidate running on a strong 'tough on crime' stance could be synthetically shown to be taking a bribe or doing drugs during polling periods – making it difficult to disprove before people voted, possibly swaying results. Deepfakes don't need everyone to believe them to influence our democracy, just a handful of people in the right electorate at the right time is all it takes to potentially change the outcome of an election.

Ukraine-Russia War

A deepfake video purporting to show the President of Ukraine, Volodymyr Zelenskyy, calling on Ukrainian troops to lay down their arms and surrender to Russian forces circulated on social media and was briefly placed on a Ukrainian news website in March 2022 before being removed.[205] It was also circulated on social media, with one version receiving more than 120,000 views on Twitter.[206]

The timing and subject matter made the video potentially significant. A fabricated surrender announcement from Ukraine's President could have created confusion, weakened morale and influenced the conduct of soldiers and civilians during an active conflict. This highlights the growing role deepfakes are having both domestically and abroad.[207]

The use of deepfakes during this conflict confirmed fears by intelligence agencies, like the FBI, about an increasing use of deepfakes against western democracies.[208] The incident demonstrated how synthetic media could be part of an information operation.

Propaganda and false claims will increase, as synthetic media lowers the cost to entry, while also expanding the reach of content shared online.[209] In recent times propaganda has been generated by individuals in places such as China's '50-centres' and Russia's 'troll farms'.[210] However, the emergence of increasingly sophisticated and inexpensive AI products may see humans removed from the process as AI-generated content is cheaper, faster and more effective in information warfare where scalability is essential.[211] The Australian Strategic Policy Institute (ASPI) claims that deepfakes will be particularly effective in the Russian model of propaganda – which focuses on volume and rapidity of false information to disorient and divide populations.[212]

Although information warfare has, until now, primarily been the domain of state actors – the availability of synthetic technology means that non-state actors will become increasingly active in this space.[213]

Finding

Synthetic media is enabling democratic interference at low cost and significant scale, while making harms difficult to prevent and reverse. Its speed, reach and potential use at critical moments make manipulation difficult to detect and counter before harm occurs. Even where synthetic content is subsequently corrected or removed, harms to individuals, public trust and democratic participation may endure.

This analysis shows that synthetic media is expanding the scale and reach of interference in ways that are difficult to detect and counter in real time. By exploiting the speed and reach of digital platforms, it allows state actors to influence democratic processes. This places pressure on democratic resilience by weakening public confidence in elections and institutions.

4. A layered response to synthetic media

This chapter assesses how Australia can prevent harm, respond when synthetic media circulates, and provide remedies and accountability after harm occurs. These three functions explain what the layered response must do. The legal, technical and civic measures examined in this chapter explain which tools and institutions can contribute to each function. The chapter identifies reforms that warrant action now, while explaining the safeguards and complementary measures needed for them to be effective. It concludes that no single measure or institution can manage the risk alone.

Finding

Measures to prevent, respond to and remedy harm must protect freedom of expression and institutional neutrality. Responses should address systemic harm without making a government authority the sole arbiter of truth or compromising the neutrality of electoral institutions. Broad prohibitions or takedown powers may capture legitimate satire and commentary, while complaints and removal pathways require clear safeguards and proportionate thresholds.

Finding

Australia's existing laws provide important but limited protection against the harms of synthetic media. Electoral and criminal laws may help prevent or deter some harms, while online safety and complaints schemes can support intervention when harmful content circulates. Defamation, copyright, privacy and vilification laws may provide remedies after harm occurs. However, these pathways are generally reactive, individually focused or difficult to enforce at the speed and scale at which synthetic content circulates.

Finding

Technical measures can support transparency, prevention and timely response, but are not a complete solution. Visible labels, embedded metadata and other authenticity measures should be standardised and interoperable so they can reduce the risk of deception, assist intervention when harmful content circulates and provide information for oversight and accountability. They cannot reliably determine what is real or prevent harmful content from circulating on their own.

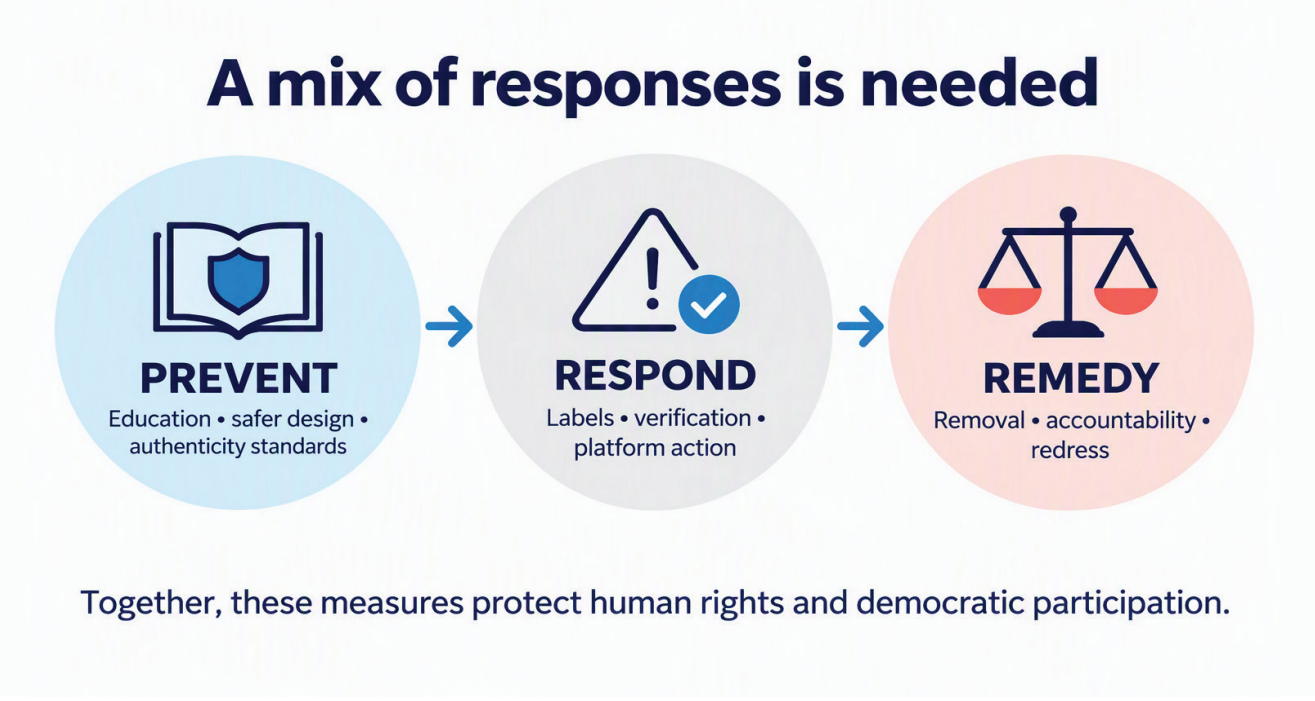
Finding

Australia requires a coordinated and layered national response that can prevent harm where possible, respond effectively when synthetic media circulates and provide remedies and accountability after harm occurs. No single measure is sufficient. These functions require complementary legal, technical and civic measures, including proportionate laws, technical transparency, digital and civics education, independent media, responsible content creators and trusted public institutions. Measures may contribute to more than one function and should reinforce one another while protecting human rights and democratic participation.

This chapter explains how a set of complementary reforms can reduce the democratic risks posed by synthetic media. Electoral safeguards, targeted legal remedies, technical transparency, education, independent media and trusted institutions each address a different point at which harm may arise. These measures should be pursued as a layered package, with their design calibrated to the function and limits of the relevant institution.

The layered response recognises that there is no single solution. Prevention, response and remedy are objectives while legal, technical and civic measures are tools that can help achieve those objectives. A single measure may play more than one role. A legal measure may deter conduct, support intervention during a critical moment or provide redress after harm. Technical and civic measures may likewise operate across more than one function. The value of the framework lies in identifying the role and limits of each measure, rather than assigning every measure to a single stage.

A layered approach reflects the reality that synthetic media harms arise from structural features of information environments rather than isolated misuse. It avoids concentrating responsibility in a single institution and instead distributes functions across actors that are suited to their roles. This structure helps ensure that failures in one area do not undermine trust in the whole system.



4.1 Electoral laws

Electoral laws play a central role in protecting the integrity of elections by setting boundaries on conduct and supporting public confidence in electoral outcomes. In Australia, this role is carried out by the regulator, the Australian Electoral Commission (AEC), whose independence is fundamental to democratic legitimacy. Electoral legislation focuses on regulating voting processes and formal campaign communication, reflecting an environment in which electoral interference was relatively contained and could be attributed to identifiable actors. Synthetic media challenges these assumptions by enabling forms of manipulation that circulate rapidly across digital platforms and frequently fall outside enforcement.

Existing electoral laws in Australia are constrained by their narrow scope and their reliance on remedies that take effect after harm has occurred. Proposals to expand laws in response to synthetic media must contend with the serious risk that expansion could weaken confidence in the neutrality of electoral institutions - particularly if it requires electoral authorities to make contested judgments about the truthfulness or legitimacy of political communications. This limits the extent to which electoral law can function as a primary mechanism for managing the potential impact of deepfakes on democratic processes and underscores the need for caution in relying on electoral regulation alone.

Commonwealth laws

The *Commonwealth Electoral Act 1918* (Cth) is the primary legislation, which makes it an offence to publish material that is likely to mislead or deceive an elector in relation to casting a vote.[214] These laws were developed for a media environment in which misleading conduct was more readily identifiable and typically linked to known sources. These laws are technology neutral and apply where synthetic media mislead voters about how to cast a vote or where and when voting occurs.

The Act does not cover synthetic media aimed at manipulating politics and democratic discourse more broadly.[215] As a result, the Act has limited application to the most harmful uses of synthetic media in democratic contexts. One possible response would be to expand electoral laws beyond their current application limited to communications about the mechanics of voting. However, doing so would require regulators to assess the truth or accuracy of political communications, raising significant practical concerns and risks of perceived bias.

The AEC occupies a unique position as the institution entrusted with administering electoral laws impartially and without political influence. Confidence in this independence is a foundation of electoral legitimacy. For this reason, regulatory responses to synthetic media must be carefully calibrated to avoid expanding the AEC's role in ways that could undermine public perceptions of neutrality. Requiring the AEC to make determinations about contested political content, including whether material is synthetic, risks drawing the AEC into controversy. This risk arises not only where the AEC fails to identify or respond to deceptive content, but also where its enforcement appears uneven or delayed. During an election period, decisions about whether, when and against whom to take action may be perceived as political interventions. Such approaches may also raise concerns under Australia's implied constitutional freedom of political communication if restrictions were not clearly proportionate to the risk being addressed.

While addressing synthetic media is necessary to protect information integrity during elections, regulatory responses must avoid positioning the AEC as an arbiter of political expression. If confidence in the impartiality of the electoral umpire is weakened, the resulting damage to democratic trust may exceed the harm caused by individual instances of deceptive content. Preserving institutional legitimacy must therefore remain a central consideration when designing electoral responses to synthetic media.

Truth in political advertising

Truth in political advertising laws have been proposed as one response to some of the electoral risks posed by synthetic media, particularly where it is used in electoral advertising to convey false information as fact. Evaluations of the longstanding South Australian scheme suggest that such laws may influence campaign practices and encourage greater scrutiny of electoral advertising.[216] Parties more closely scrutinise advertisements, compliance with directions to withdraw or retract material is generally reported to be strong, and the reputational consequences of an adverse finding may operate as a practical deterrent.[217] Participants in South Australia did not report a chilling effect on political debate. This may reflect the scheme's narrow focus on verifiable statements of fact rather than opinions or predictions,[218] a limitation that may assist in maintaining compatibility with the implied freedom of political communication in the Australian Constitution.

However, truth in political advertising laws have important limitations as a response to synthetic media. A narrow scheme may be more constitutionally and administratively workable, but it will not capture many forms of synthetic media that convey political messages through implication, opinion, prediction, satire or other forms of expression that do not readily lend themselves to objective verification.[219] There are also practical difficulties in resolving complaints quickly during election periods, securing equivalent prominence for retractions, and addressing anonymous, overseas or rapidly disseminated synthetic content. More broadly, expanding the scope of such laws raises difficult questions about who should determine the truth or falsity of political claims and how any such framework can operate without unduly burdening political communication or undermining the institutional neutrality that is an essential feature of Australia's electoral regulators. These considerations suggest that truth in advertising laws, if adopted, should be regarded as only one possible element of a broader response to the risks posed by synthetic media.

Prohibiting political deepfakes

The *Electoral Legislation Amendment (Electoral Communications) Bill 2024* (Cth) was introduced to prohibit the dissemination of certain inaccurate and misleading electoral matter during election and referendum periods.[220] The Bill proposed to directly prohibit the dissemination of deepfakes during elections by a person or entity authorised to make communications about electoral matters.[221] The Bill proposed the creation of an Electoral Communications Panel, responsible for monitoring for contraventions, conducting investigations and making decisions – which could lead to civil penalties.[222] The Bill lapsed when the Parliament was dissolved before the 2025 federal election.

Electoral laws at the state and territory level have made greater attempts to address political deepfakes. Tasmanian and Australian Capital Territory laws prohibit the dissemination of an advertisement that contains inaccurate or misleading statements of fact.[223] These provisions may extend to synthetic media which is framed as electoral advertising and conveys misleading statements of fact. While no case about synthetic media has been heard in these jurisdictions, it is likely that future cases may be decided in coming years.

South Australia has laws that specifically address some of the issues raised by deepfakes. As of 2025 the *Electoral Act 1985* (SA) was amended to expressly prohibit electoral advertising that depicts a real person doing or saying something they did not do or say without that person's permission.[224] A breach of the provision carries a penalty of $5,000 for a person and $10,000 for a body corporate.[225] Additionally, such breaches provide both the Electoral Commissioner and the Supreme Court powers to request the advertisement be withdrawn and/or publish a statement relating to the contravention.[226] The Court of Disputed Returns could also invalidate an election where misleading advertising affected the result.[227]

This aligns with a growing trend overseas seeking to ban deepfakes during elections. In late 2023 South Korea amended its Public Official Election Act to ban the creation and posting of deepfakes for electoral purposes in the 90 days leading up to its 2024 general election.[228] Equally in Singapore a prohibition on deepfakes targeting political candidates was passed in 2024.[229] Meanwhile in Taiwan, laws were passed in 2023 that allow candidates to report deepfakes of themselves to social media companies for removal.[230]

Prohibitions on deepfakes during election campaigns have serious practical constraints in application. As deepfakes become increasingly difficult to distinguish from authentic material, regulators cannot reasonably be expected to readily identify and verify synthetic media at the pace required during election periods. Technical measures that rely on detection or labelling remain limited in effectiveness and cannot provide consistent coverage in fast-moving information environments.

Regulatory failure to promptly identify and respond to specific deepfakes can further undermine public trust. When manipulated content circulates without intervention, or is addressed inconsistently, perceptions of neutrality and competence in electoral oversight may be weakened.

Synthetic content during the Secret Harbour by-election

During the August 2026 Secret Harbour by-election in Western Australia, synthetic content was used as part of the online campaign against Labor candidate Georgia Tree.[231] Media reports identified a deepfake involving Tree among the social media material circulating during an increasingly intensive and targeted campaign for the seat.[232]

Following Labor's loss of the seat, federal Resources Minister Madeleine King raised concerns about the role of social media bots and misinformation.[233]

The case illustrates the limitations of relying on electoral remedies after synthetic content has circulated. Even where content is identified and corrected, a tightly timed response may be difficult during a short by-election campaign. It also reinforces the need for electoral authorities, political parties, candidates and platforms to establish clear processes for identifying and responding to synthetic political content across all levels of government.

Authorisations

The South Australian electoral laws also include an obligation to label any AI-generated electoral advertisements.[234] A breach of the provision carries a penalty of $5,000 for a person and $10,000 for a body corporate.[235] Additionally, a breach also allows both the Electoral Commissioner and the Supreme Court powers to request the advertisement be withdrawn and/or publish a statement relating to the contravention.[236]

Authorisation and labelling requirements should be directed specifically at deepfakes that falsely represent real people or events. Content of this kind has a distinct capacity to mislead in political contexts because it presents fabricated material as authentic. By contrast, AI-assisted production is now embedded across a wide range of communication practices. Extending regulation to all forms of AI-generated content risks over-capture and dilution of enforcement, while failing to focus on the content most likely to undermine democratic processes.

Authorisation requirements can support transparency by assisting voters to understand the source of formal campaign material. However, the most harmful uses of deepfakes during election periods are likely to originate from actors who operate outside authorisation regimes. These actors may not be subject to disclosure obligations and can disseminate manipulative content without attribution. Authorisation therefore remains a useful but limited safeguard that cannot address the primary sources of deepfake-driven harm.

These laws can improve transparency and accountability but cannot fully reverse the effect of a well-timed deepfake on an election outcome, particularly once votes have been cast.[237] Ongoing injunctions restraining publication or republication may assist in few cases, provided an identifiable person or party is within Australia's jurisdictional reach.[238]

It would potentially only take a few people to be influenced by a deepfake to change the outcome of an election.[239] Just a few votes can change the course of an election. For example, the seat of Bradfield was won in 2025 by just 26 votes.[240] Concerningly broader research by McAffee has shown that 48% of respondents felt deepfakes could influence who they voted for.[241]

Electoral laws should form a targeted part of the layered response. While noting important limitations, authorisation and labelling requirements can help prevent deception by reinforcing trust in formal political communication. Clear election-period processes can also support a timely response by improving coordination between electoral authorities, political participants and platforms. Their role should remain focused so that the independence of electoral institutions is protected, while complementary legal, technical and civic measures address deceptive content that falls outside formal campaign regulation or circulates too late for an electoral remedy to prevent harm.

4.2 Deepfake sexual abuse material laws

Deepfake sexual abuse material raises distinct concerns because it involves the non-consensual sexualisation of people through digital means. While the immediate impact is often deeply personal, this conduct also limits democratic participation by increasing the personal risk associated with public visibility.

This section examines how existing legal frameworks respond to these harms, with particular attention to criminal and online safety laws.

Criminal laws

Criminal laws play an important role in responding to deepfake sexual abuse which can be used to intimidate or humiliate women who engage in public life. This form of abuse interferes with rights to privacy and dignity, while also reshaping the conditions of democratic participation by attaching heightened personal risk to public visibility. While recent reforms address distribution, criminal laws largely operate after material has been shared, rather than recognising the creation of deepfake sexual abuse as a harm.

In 2024, the Australian Government passed the *Criminal Code Amendment (Deepfake Sexual Material) Act* (Cth) which created new federal offences for sharing sexually explicit material without consent. This includes deepfake content that appears to show a real individual.[242] The offence applies when a person knowingly or recklessly distributes such material, with penalties of up to seven years' imprisonment.[243]

Additional criminal offences may apply where synthetic content targets children and young people, including under state and territory laws. However, these protections are not nationally consistent and remain subject to the enforcement limitations. They also operate primarily as responses to individual conduct, rather than addressing the systemic conditions that enable harmful synthetic content to be created and distributed.

These offences depend on the ability of investigators and courts to determine whether material is synthetic or genuine, a distinction that is increasingly difficult to establish in practice. Advances in generative technology allow deepfake material to closely resemble genuine imagery, complicating criminal investigations. These challenges are particularly acute where deepfake sexual abuse material involves young people, including peer-to-peer abuse among students, which may engage legal regimes relating to child sexual abuse material.[244] In these situations, uncertainty about the nature of the material does not reduce the harm experienced by victim-survivors, but it can delay protective action and hinder effective accountability.

Notably it was a 19-year-old-boy who, in April 2026, was the first to be prosecuted under the new laws – pleading guilty to two counts of creating or altering sexual material without consent and two counts of using a carriage service in a harassing or offensive way.[245]

However, the Act did not create a standalone offence for the creation of deepfake sexual abuse - instead, criminal liability depends on the sharing of material.[246] Establishing a standalone offence would recognise the violation experienced by victims (including those who participate in public life) in the act of creating deepfake sexual abuse material. This is important because the deterrent effect on female participation in democracy does not depend solely on dissemination. The knowledge that public engagement may prompt the private creation of deepfakes sexual abuse material may be sufficient to discourage participation. Recognising creation as an offence would address this gap by affirming that non-consensual deepfake creation is itself a serious violation that can function to silence women.

Online safety framework

Deepfake facilitated abuse raises distinct concerns because it often involves the non-consensual sexualisation of people through digital means. This form of abuse can target anyone, although women are disproportionately targeted.

Australia's online safety framework provides important tools for responding to technology-facilitated gender-based violence (TFGBV), but it was not designed to address the democratic risks posed by synthetic media. The *Online Safety Act 2021* (Cth) (OS Act) does not regulate synthetic media where it undermines democratic engagement as such. Instead, it regulates defined categories of online harm, some of which may incidentally capture certain deepfakes (including image abuse or adult cyber abuse). As a result, the framework offers partial protection in democratic contexts but leaves gaps where deepfakes affect participation without meeting established content classifications.

The Image-Based Abuse Scheme (IBAS) provides eSafety with powers to remove non-consensual sharing of, or threat to share, an intimate image online and/or impose a civil penalty.[247] The Commissioner can also issue a removal notice to an internet service that is hosting the image.[248] Notably the IBAS applies to deepfakes.[249] For example, a deepfake image of a political candidate having a sexual affair, released just before an election, would fall within the definition of 'intimate image'.

Where a person believes an intimate image of them has been shared without their consent, they can make a complaint to eSafety. In the first instance, eSafety typically approaches online service providers to ask them to remove an intimate image, or to alert them to an account which is threatening to post the image.[250] Although this can be a timelier solution it has been criticised because it does not hold perpetrators to account.[251]

If more formal action is preferred, eSafety can issue removal notices to both the end user and online service provider to take all reasonable steps to remove the material within 24 hours.[252] This is crucial to minimising the harm of deepfake sexual abuse material.

Should a removal notice not be complied with, eSafety can:

- issue a formal warning
- accept an enforceable undertaking
- seek a court injunction
- issue an infringement notice
- seek a civil penalty order.[253]

This can result in significant financial penalties for individuals or online service providers. In a landmark case, a defendant was fined approximately $340,000 for creating and posting deepfake sexual abuse material of six prominent Australia women.[254]

As with other complaints-based schemes, the IBAS depends on affected people being aware that the deepfake exists in order to bring it to the regulator's attention. In an information environment characterised by volume and fragmentation, this may not always be possible. Deepfake sexual abuse material may circulate within closed or anonymised networks, leaving those targeted unaware that material has been created. In such circumstances, harm may persist without any regulatory response. The possibility of unseen circulation may itself deter political engagement, even where no complaint is made. The IBAS therefore reflects a constrained approach that responds to identified harm.

Part 7 of the OS Act addresses cyber abuse directed at Australian adults. Cyber abuse material is content that an ordinary reasonable person would regard as intended to cause serious harm and as having a menacing or harassing character.[255] Where this threshold is met, eSafety may issue removal notices to individuals or hosting services, with civil penalties available for non-compliance.[256]

The application of this framework requires particular care in relation to politicians, candidates and others engaged in public debate on matters of public interest. Democratic participation depends on tolerance for robust criticism and scrutiny of public figures. However, this does not diminish the importance of protecting individuals from threatening, harassing or otherwise seriously harmful synthetic media. This distinction is reflected in some platform practices that adopt a higher intervention threshold for public figures in order to safeguard political communication.[257]

Deepfakes may fall within the cyber abuse framework where they are used to intimidate or threaten. However, application must remain sensitive to democratic context. Over-extension risks treating political expression as abuse, while under-application risks leaving harmful impersonation without remedy. This tension highlights the limits of applying general online safety tools to harms that arise from political participation.

The online safety framework generally responds to identifiable incidents affecting particular individuals. This can make it difficult to recognise and respond to coordinated or high-volume attacks, where each item of content is treated separately despite forming part of a broader campaign of abuse.

Collective Shout: deepfake sexual abuse and the online safety framework

Following Collective Shout's 2025 campaign against online games containing depictions of rape and other forms of sexual violence against women, members of the organisation were subjected to a sustained campaign of technology-facilitated abuse.[258] The Online Hate Prevention Institute documented a sustained campaign of image-based abuse, including deepfake pornographic material, manipulated sexual imagery and other AI-generated content depicting Collective Shout staff in sexualised, degrading and violent scenarios, as part of a broader campaign that also involved doxxing, harassment, and rape and death threats.[259] The report found that this material was used not only to humiliate individual women, but also to discourage and constrain their participation in public debate by sexualising and subordinating women who had exercised public influence.[260]

The case highlights limitations in Australia's online safety framework when responding to coordinated technology-facilitated abuse campaigns. The framework depends heavily on affected people identifying and reporting particular items. Collective Shout submitted reports to platforms and sought eSafety's assistance, yet the report records that some platform complaints were rejected, harmful material remained available, and the reporting process was resource intensive.[261] It illustrates how an incident-based model can provide a pathway for individual removal while struggling with coordinated, high-volume and cross-platform campaigns.

The online safety framework provides an important remedial layer by enabling removal and enforcement in some individual cases. However, because they depend on identified content and affected individuals bringing harm to the regulator's attention, they should operate alongside preventive education and product design, as well as systemic responses to coordinated amplification and high-volume abuse. Combining these functions would better address both the immediate needs of people targeted by deepfakes and the broader participation-limiting effects of synthetic media.

At the time of publication the Online Safety Amendment (Digital Duty of Care) Bill 2026 is being considered by Parliament. Although the Bill contains targeted measures relating to fake nude material and may address some deepfake-related harms through its broader duty-of-care framework, it is too early to assess its final form and effect.

4.3 Misinformation and disinformation laws

Synthetic media also raises wider issues relating to misinformation and disinformation. In 2023, the Australian Government released the Exposure Draft of the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* (Cth), aiming to regulate digital platforms through oversight by the Australian Communications and Media Authority. While the intent was to curb harmful content online, the Commission raised serious concerns about the draft's lack of safeguards for free expression and its potential to enable censorship.[262]

Although the Exposure Bill was revised and reintroduced in 2024 with some improvements, the proposed laws still failed to appropriately balance protecting freedom of expression with addressing misinformation and disinformation.[263] Ultimately, the legislation was not pursued.[264] In democratic systems, disputes about facts are often inseparable from political contestation – this makes drawing the line about which content to moderate or not difficult.

In the absence of binding legal requirements, misinformation and disinformation in Australia is currently primarily addressed through voluntary platform governance as developed by the Digital Industry Group Inc.[265] For example, TikTok guidelines state that the platform does not allow 'misinformation that could prevent people from voting, interfere with elections, or encourage the unlawful disruption of results.'[266]

TikTok has also taken further measures against misinformation in relation to elections. During election times, the platform uses a combination of technology and human moderators as well as a Global Fact-Checking Program to assess content, label content that has not been verified and removing or reducing the reach of misinformation. These fact-checking organisations contribute to identifying misinformation that the app then moderates directly.[267]

Other platforms adopt different approaches. YouTube prohibits certain content that misleads people about voting, advances false claims about a candidate's legal eligibility or encourages interference with democratic processes. Its policy also addresses some technically manipulated content that creates a serious risk of harm, while allowing exceptions for educational, documentary, scientific, artistic or public-interest content.[268]

These examples demonstrate the range of measures available to platforms, but their effectiveness and continued application depend on the policies and enforcement decisions of each company.

As observed by the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, platforms exercise substantial influence over public discourse despite lacking democratic mandates.[269] They are not democratic institutions and although voluntary measures should play a role, platforms cannot have a sole determinative role in assessing truth within democratic debate.

Finding

Measures to prevent, respond to and remedy harm must protect freedom of expression and institutional neutrality. Responses should address systemic harm without making a government authority the sole arbiter of truth or compromising the neutrality of electoral institutions. Broad prohibitions or takedown powers may capture legitimate satire and commentary, while complaints and removal pathways require clear safeguards and proportionate thresholds.

4.4 Other areas of law

Reactive legal pathways allow individuals to seek redress after deepfake harm has already occurred. They rely on existing bodies of law that were not developed with synthetic media in mind. In democratic contexts, reliance on reactive pathways has clear limits. Remedies that depend on litigation are slow and costly - rarely mitigating the immediate participation-limiting effects of synthetic media. They also tend to treat each instance of harm separately, making it difficult to recognise and respond to coordinated, high-volume attacks or their cumulative effects.

Copyright and intellectual property

Copyright and intellectual property law offer reactive pathways through which people may seek to challenge deepfakes after harm has occurred. These legal frameworks were developed to protect proprietary interests in creative works and personal likeness, rather than to address synthetic impersonation or democratic harm.

For example, copyright laws have already been successfully used in the US by Kim Kardashian to remove a deepfake from YouTube.[270] This was done by using YouTube's existing copyright infringement procedures because the deepfake used footage owned by Conde Nast. Unfortunately such claims are costly and time consuming for the plaintiff – assuming the creator of the deepfake is even identifiable (as most are posted anonymously) or within jurisdictional reach.[271] While established political figures may have the resources to pursue copyright-based takedown action, people seeking to enter public life are far less likely to be able to rely on these pathways in practice.

Denmark has proposed amending its copyright law to protect people against realistic, digitally generated imitations of their appearance and voice. The new copyright laws being considered would grant people the right to challenge and remove unauthorised digital reproductions of their likeness or voice. It would apply to all individuals, not only public figures, and would include exceptions intended to protect parody, satire, caricature and social criticism. [272] The proposed legislation intends to strengthen copyright laws and will treat 'likeness' as a copy-right style entitlement. This will enable individuals to issue takedown notices and claim compensation.[273] However, questions remain about enforcement against anonymous or overseas actors, interaction with existing privacy and platform laws, and how exceptions for legitimate expression would operate in practice.

In the United States there have been moves to protect intellectual property rights in the voice and visual likeness of individuals – which would apply to deepfakes.[274] The federal No Fakes Act of 2025 is a bill that would give everyone proprietary rights in their image and voice and prohibits the use of deepfakes of people without a licence. Under the bill an unauthorised digital replication could result in fines of:

- $5,000 USD per deepfake for an individual
- $25,000 USD per deepfake for a provider of an online service that has made good faith efforts to comply with the bill
- Between $5,000 and $750,000 USD for a provider of an online service that has not undertaken a good faith effort to comply with the bill.[275]

The Senate bill has not yet been passed by Congress or enacted into law.

Copyright and intellectual property pathways face practical limits that reduce their usefulness in responding to deepfakes. The costs of legal action and the need to identify a responsible defendant make these pathways inaccessible for most people.

Defamation

People who find themselves targeted by a deepfake may have a cause of action under defamation law – provided a plaintiff can prove the material was published by the defendant, it identified the plaintiff (either directly or indirectly), and that the material is defamatory.[276] Unfortunately, defamation litigation is also slow, expensive and unlikely to repair the damage done by a deepfake. Similarly to copyright, it also requires the identification of the creator of the deepfake and for them to be within jurisdictional reach.

In May 2024, an independent candidate for Mayor of the West Midlands in the UK, shared a deepfake on TikTok depicting a Labour Party researcher using racist language.[277] Although the mayoral candidate did not create the deepfake, his decision to share it with his large following triggered public abuse and threats against the researcher.[278] Action was commenced under the United Kingdom's *Defamation Act 2013* (UK) which resulted in a substantial outside of court settlement. Notably although the candidate did not create the deepfake they were the target of the legal action for simply resharing the video. This was the first time a deepfake in the UK had been litigated in the courts for defamation.

Defamation law operates retrospectively and is poorly suited to address deepfakes. Its dependence on identifying a publisher and its slow operation mean it cannot prevent reputational or participatory harm once synthetic content has circulated. As a result, it offers limited protection where deepfakes deter engagement before legal action can realistically begin.

Hate speech and vilification laws

Synthetic media may engage federal racial vilification law where it publicly depicts people of a particular race, colour or national or ethnic origin in a demeaning or threatening way. Section 18C *Racial Discrimination Act 1975* (Cth) makes an act unlawful where it is done publicly, is reasonably likely to offend, insult, humiliate or intimidate a person or group, and is done because of their race, colour or national or ethnic origin.

Whether synthetic content contravenes s 18C will depend on its content and context, including how it depicts the relevant group, the language accompanying it and the audience to whom it is communicated. The provision may apply even where the synthetic content does not identify a particular individual, provided it targets a racial or ethnic group. Section 18D provides exemptions for certain acts done reasonably and in good faith, including artistic works, reporting and fair comment on matters of public interest.

However this process generally depends on identifying a respondent, which may be difficult where synthetic content is created anonymously, circulated from outside Australia or repeatedly reposted.

Privacy

Privacy law offers another avenue for potential remedy where a person's image or voice is misused through deepfakes. Recent reforms signal a growing recognition of digital privacy harms, which may be relevant to deepfakes.

The *Privacy Act 1988* (Cth) includes a statutory tort for serious invasions of privacy that are intentional or reckless.[279] The tort is confined to invasions either by intrusion upon seclusion or misuse of private information.[280] The tort was informed by the Australian Law Reform Commission's (ALRC) Report 123, which considered how Australian law should be reformed to prevent and remedy serous invasions of privacy.

The essential elements of the tort are:

- the invasion of privacy must be either by intrusion into seclusion or by misuse of private information
- it must be proved that a person in the position of the plaintiff would have had a reasonable expectation of privacy in all of the circumstances
- the invasion must have been committed intentionally or recklessly – mere negligence is not sufficient
- the invasion must be serious
- the invasion need not cause actual damage, and damages for emotional distress may be awarded
- the court must be satisfied that the public interest in privacy outweighs any countervailing public interests.[281]

It is currently judicially untested if the creation of deepfake would allow a plaintiff to bring a cause of action under the tort.

In November 2025 Senator David Pocock introduced the *Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025* (Cth) (My Face, My Rights Bill) as a private Senator's bill. The Bill sought to use existing legal frameworks to affirm that 'every Australian retains ownership of their image, voice, and identity in the digital world'.[282]

Schedule 1 of the Bill proposed a complaints system (managed by eSafety) for people to flag deepfakes depicting them.[283] This could then lead to investigation and enforcement action by eSafety.[284] Schedule 2 of the Bill introduced a civil cause of action under the Privacy Act for wrongful use or disclosure of deepfake material. There were exemptions to cover journalists who publish a deepfake as part of journalistic material among other exemptions.[285]

Remedies available under the proposed Bill included:

- injunctions
- orders requiring the defendant to apologise
- correction orders
- orders requiring the defendant to destroy certain materials connected to the deepfakes, and
- damages.[286]

The My Face, My Rights Bill's emphasis on individual control over image and identity has particular relevance to political figures who are targeted by deepfakes. Frameworks of this type may offer one means of addressing impersonation and reputational harm without requiring electoral regulators to determine the truth or falsity of contested political claims. However, any expansion of legal remedies in this area would need to be carefully balanced against the importance of protecting freedom of expression, including political commentary, satire, parody and other forms of legitimate public discourse. As with other complaints-based schemes, effectiveness would depend on awareness, accessibility and timely action. In the context of political deepfakes, harm often occurs at the point of circulation, and remedies that operate primarily after dissemination may provide only limited protection for democratic integrity. These considerations illustrate both the potential utility and the limitations of individual complaints-based approaches as part of a broader response to synthetic media.

International experience shows that impersonation laws can suppress lawful expression when safeguards are weak. In jurisdictions such as India, broad takedown powers linked to impersonation have been used against political criticism and satire. This highlights the need for clear limits that distinguish harmful impersonation from legitimate speech - preventing reactive remedies from becoming tools of censorship.

Xi Jinping and Winnie the Pooh

In 2013, internet users began comparing Chinese President Xi Jinping with Winnie the Pooh after images of Xi Jinping walking alongside then US President Barack Obama were likened to the fictional characters Winnie the Pooh and Tigger.[287] The comparison developed into a widely recognised form of political satire, with further memes using the character to comment on Xi's public image and leadership. Chinese authorities and online platforms subsequently restricted some images and references connecting Xi with Winnie the Pooh after the comparison became associated with political criticism and symbolic dissent.[288]

The satire has since expanded beyond conventional memes to include AI-generated images depicting Xi interacting with Winnie the Pooh. These images demonstrate how synthetic media can provide new forms of political parody and criticism by adapting an established cultural reference.

The example illustrates the risk that broadly framed impersonation or takedown powers may capture other forms of expression as well as harmful deepfakes. A removal scheme designed to give individuals control over the use of their likeness could be used by political figures to suppress satire or dissent if it does not contain clear safeguards.

Defamation, copyright, privacy and related private-law pathways can help people seek remedies for harms caused by synthetic media. They may provide removal, correction, compensation or accountability in individual cases after a deepfake has been circulated. Because these pathways rely on individual action, identifiable respondents and processes that may be slow or costly, they are sufficient on their own. People also need accessible complaints processes and effective measures to prevent and respond to harmful deepfakes before private legal action can realistically take effect.

Finding

Australia's existing laws provide important but limited protection against the harms of synthetic media. Electoral and criminal laws may help prevent or deter some harms, while online safety and complaints schemes can support intervention when harmful content circulates. Defamation, copyright, privacy and vilification laws may provide remedies after harm occurs. However, these pathways are generally reactive, individually focused or difficult to enforce at the speed and scale at which synthetic content circulates.



4.5 Technical responses

This section examines technical solutions to synthetic media which undermines democratic participation and resilience. Technical measures can assist people to recognise manipulated content and can support regulatory decision-making by improving transparency. However, technical tools also have limits. This section therefore considers labelling and platform architecture as supporting responses that operate within a broader framework.

Labelling

To assist people in identifying synthetic media, the Australian Government is considering the introduction of voluntary watermarking for AI-generated content.[289] Watermarks can serve as a practical tool to help users identify synthetic content, improving transparency and reducing the risk of deception.[290]

As synthetic media becomes harder to detect, labelling will play a role in responding to synthetic media. Other countries are already ahead of Australia in this regard.

The European Union regulates synthetic media transparency through the EU AI Act[291] and the Digital Services Act,[292] which together impose obligations on those who create and distribute synthetic content. Article 50 of the EU AI Act[293] requires providers of AI systems that generate synthetic media to ensure outputs are marked in a machine-readable form that identifies them as generated, while professional deployers must disclose to audiences when content has been synthetically created or manipulated.[294]

China has adopted a highly prescriptive approach to labelling through binding rules and technical standards that apply across the digital content environment.[295] The framework requires synthetic content to carry a visible label presented directly on the content, alongside embedded metadata identifying the content as synthetic and recording its source – ensuring transparency both for viewers and within the underlying file.[296] These obligations extend to platforms who are required to apply labels where content is identified, or reasonably suspected, to be AI-generated even in the absence of prior disclosure.[297] By mandating labelling at the point of viewing and within metadata, China's approach treats transparency as an enforceable structural feature that supports detection and accountability throughout the content lifecycle.

Metadata-level labelling is critical for effective regulation because many legal and administrative responses depend on determining whether content is synthetic. Regulators such as eSafety need reliable ways to distinguish deepfake material from authentic imagery so that the correct scheme is applied and misclassification is avoided. This distinction is especially important where different legal consequences attach to deepfake-specific offences, including criminal laws relating to synthetic sexual abuse material, as compared to offences involving real non-consensual imagery. Consistent metadata identifying content as synthetic reduces reliance on subjective visual assessment or disputed claims by content creators, supporting more reliable regulatory decision-making.

Labelling content without reliable evidence that it is synthetic may restrict legitimate expression and cause confusion. It may also lead people to assume that any content without a label is authentic. This could create a false sense of confidence in unlabelled content, making it easier for synthetic content to be mistaken for authentic content. At the same time, inconsistent or inaccurate labelling may undermine confidence in genuine content by encouraging people to question its authenticity. Labelling should therefore be required only where there is a reliable basis for determining that content is synthetic.

Proudly human

Proudly Human is a voluntary certification scheme that verifies and labels content created by humans rather than generated by AI.[298]

It operates as a human-origin disclosure mechanism, allowing creators and organisations to apply a certification mark that signals to audiences that a work is human-created following an independent verification process. Proudly Human is not a detection system and does not regulate the use of AI. Articles are already appearing with a Proudly Human certification in Australia.[299]

Its function is to support informed choice in information environments where AI-generated material is increasingly difficult to distinguish from human expression. In the context of deepfake labelling, Proudly Human demonstrates how voluntary certification can complement regulatory approaches by signalling human authorship, while also underscoring the limits of opt-in models where participation cannot be assumed.

Industry collaborations such as the Content Authenticity Initiative are already seeking to address misleading synthetic media by adding digital watermarks to identify whether visual content is real or synthetic.

TikTok has policies for content about elections and civic participation. AI-generated content showing realistic people or events must be labelled. If it is not labelled, TikTok may label, restrict or remove it.[300] It also bans a range of AI-generated content, such as content that pretends to be a real news source, or deepfakes of a public figure taking political stances or endorsements that did not really occur.[301]

Labelling provides an important role in improving the integrity of information online, however it is not perfect nor a panacea. For example, there is no standardised labelling technology/procedure in Australia – meaning that one system may not be interoperable with another.[302] Even where labelling is required, research is indicating that:

- metadata watermarking – which embeds the mark directly in the metadata – is vulnerable to manipulation, and all watermarking methods are vulnerable to manipulation at some level. [303]
- the technical difficulty in detecting deepfakes, including the challenge of training an algorithm to do this effectively, requiring a large amount of labelled deepfakes which is difficult to obtain.[304]
- lack of consistent or practical guidance as to how entities are to comply with their regulatory requirements. [305]

For labelling to strengthen information integrity, labels must be accurate and applied in a consistent way. Incorrect or contested labels risk weakening public confidence in the labelling system itself. Where trust in labels erodes, people may discount them entirely, increasing scepticism toward all content regardless of origin. In these conditions, labelling can deepen uncertainty rather than alleviating it.

Further, the actual effectiveness of labelling raises concern.[306] Research indicates that labels help some people understand they are consuming generated content, but not everyone, and labels are unlikely to affect the person's engagement with synthetic media (e.g. sharing it online).[307] Labelling synthetic media with additional context may, however, reduce the number of users who believe and share false information.[308] The larger and more assertive these labels are – the more effective they can be.[309]

Labelling is an important technical response to the risks posed by synthetic media, but it cannot function as a standalone solution. Comparative experience shows that its primary value lies in enabling transparency, particularly where visible labels are reinforced by embedded metadata. At the same time, limits on effectiveness and risks to public trust mean that labelling must be narrowly and carefully applied. When used in this way, labelling can strengthen the information environment without becoming a proxy for further distortion.

Platform architecture

Platform architecture plays a significant role in shaping how information is encountered and evaluated online.

Platforms built around short-form video and continuous recommendation feeds, such as TikTok or Instagram, are designed to prioritise discovery and rapid consumption. Content is largely encountered through algorithmic curation rather than through intentional searching or prior affiliation with a defined group.[310] This design can allow engaging or emotive material, including misleading synthetic media, to reach large audiences quickly. At the same time, this architecture enables centralised intervention through measures such as automated moderation and adjustments to recommendation systems that reduce amplification during elections or other high-risk periods. In this model, responsibility for maintaining information integrity rests primarily with platform governance and technical controls.

By contrast, platforms organised around topic-based communities and pseudonymous participation, such as Reddit, create different conditions for how information is assessed and challenged. As set out in Reddit's December 2025 Transparency Report:

> Content moderation on Reddit happens through a layered, community-driven approach. This approach is akin to a democracy, wherein everyone has the ability to vote and self-organize, follow a set of common rules, establish community-specific norms, and ultimately share some responsibility for how the platform works.[311]

On Reddit content appears within defined communities that develop shared expectations about relevance and credibility.[312] Discussion is structured around comment threads, allowing claims to be interrogated publicly and in context.[313] Voting mechanisms play a direct role in shaping prominence, as comments that attract sufficient downvotes are automatically collapsed or hidden, reducing their visibility without removing them altogether.[314] This allows misleading synthetic contributions to be marginalised through collective judgement rather than immediate platform intervention.

Community moderators further shape these dynamics through their role in enforcing community-specific rules. These moderators are volunteer users with sustained engagement in their communities and familiarity with the subject matter. They can remove content that does not meet community standards and apply contextual judgement when evaluating whether synthetic or misleading material is acceptable within the forum. While the platform retains responsibility for enforcing site-wide rules, this decentralised governance structure allows scepticism and scrutiny to operate close to the point of interaction, often limiting the impact of synthetic media through discussion rather than takedown.

These differences illustrate how platform architecture produces distinct trade-offs for information integrity and participation. Algorithmically driven platforms can operate at scale and adjust systems quickly in response to emerging risks, but they may limit opportunities for slower engagement and contextual evaluation. Community-centred platforms support interrogation and peer-based challenge within discussion spaces, but outcomes can vary between forums.

Considered together, these models show that improving information integrity depends not only on moderation decisions, but also on how platforms are structured and how users are enabled to participate in evaluating the material they encounter.

Technical measures contribute primarily to prevention and response. Labels, metadata and authenticity standards can reduce the risk of deception before content is widely circulated, while platform governance and architecture determine how quickly harmful amplification can be limited when risks emerge. These tools also support accountability by providing information to regulators and researchers. Technical measures should improve transparency and enable proportionate action - without being treated as conclusive proof of authenticity or displacing rights-respecting legal safeguards.

Finding

Technical measures can support transparency, prevention and timely response, but are not a complete solution. Visible labels, embedded metadata and other authenticity measures should be standardised and interoperable so they can reduce the risk of deception, assist intervention when harmful content circulates and provide information for oversight and accountability. They cannot reliably determine what is real or prevent harmful content from circulating on their own.

4.6 Education

Education functions as a civic response to synthetic media because it affects whether people can participate in public life with confidence.

As synthetic media becomes more realistic and widespread, distinguishing truth from fabrication becomes harder.[315] This leaves people to navigate a complex and misleading digital landscape on their own. Even well-informed individuals may struggle to tell the difference between authentic and synthetic media.[316]

People who lack digital and media literacy skills are more vulnerable to deception, making them less able to participate confidently in public discourse. As a result, the digital divide is not just a matter of access – it is a driver of distrust that threatens democratic engagement and human rights. UNESCO has identified education as a core pillar in responding to harms associated with deepfakes.[317]

The digital divide is the gap between people who have access to modern technologies and those who don't. Digital inequality is evident in many ways between groups:

- urban areas versus rural communities
- rich versus poor
- educated versus under educated
- young and old.

For example, studies have demonstrated that while older people are more likely to share misinformation and disinformation – young people are more likely to believe the misleading narratives.[318]

A major challenge in countering the spread of deepfakes is the lack of access to digital literacy education. Digital literacy can significantly improve a person's ability to detect synthetic media.[319] Media and digital literacy, combined with critical thinking skills, are widely recognised as essential tools for building resilience against misleading information.[320]

Digital media literacy can help people to identify false information and unreliable news sources in the form of 'inoculation' where users are pre-emptively exposed and familiarised with the telltale signs of false information.[321] This kind of inoculation against misleading synthetic media would expose people to examples which can help them recognise and resist it in the future. This approach has been successfully applied in Australia through initiatives like *First Draft's CrossCheck Australia: Election Watch*, providing journalists with daily alerts during the 2022 federal election. By identifying data voids and anticipating public information needs, the project helped strengthen the integrity of election reporting and public discourse.[322]

Other initiatives such as the Australian Electoral Commission's (AEC) Stop and Consider campaign can also play an important role in providing tools for voters to address false information.

Public support for these strategies is strong. Surveys show that 72% of Australians believe most people struggle to recognise misleading or fake information during elections,[323] and 80% of adults want action to address the spread of misinformation.[324] Importantly, 94% of those who support action agree that education is key, underscoring the urgent need to equip Australians with the skills to navigate an increasingly complex digital environment.[325]



Civics education

Civics education is important in helping people understand how political communication now operates within digital environments. This includes awareness that some content is intentionally misleading and that democratic debate often involves disagreement. Without this context, synthetic media can distort expectations about truth in public discourse, leading to distrust toward institutions and those who take part in political debate. Civics education supports democratic culture by reinforcing shared understandings that allow participation to continue despite contested information.

Civics education is particularly important for children and young people. Although children cannot vote, they have a right to participate in public life and will become future voters. Learning how democratic institutions operate, why access to reliable information matters and how people can participate beyond voting provides an important foundation for informed democratic engagement.

Children and young people may also be particularly exposed to manipulation through the digital environments in which they communicate. Civics education should therefore be connected with digital and media literacy. While schools have an important role, this education should also occur through families, communities, public institutions and online services.

Education is a core preventive layer of the response. Digital and media literacy can help people assess synthetic content, while civics education can strengthen understanding of democratic institutions, political communication and legitimate disagreement. Education is most effective when legal and technical measures also address harmful creation and amplification, and when trusted institutions and media provide reliable information during periods of uncertainty.

4.7 Independent news and media

Independent news media remain an important civic safeguard against the risks posed by synthetic media. Professional journalism provides established processes for verifying content, although these processes require time, resources and technical capacity. Sustained support for independent media can help preserve trusted sources of information and public confidence in reporting.

New media creators also play an increasingly important role in public discourse. Constructive engagement can help responsible creators access reliable information, correct errors and distinguish clearly between reporting and commentary. Supporting these practices can strengthen credibility without imposing inappropriate gatekeeping or relying exclusively on legacy media.

4.8 Trusted institutions

Trusted public institutions help maintain a shared factual basis for democratic participation when synthetic media creates uncertainty about what information is reliable. Bodies such as the Australian Electoral Commission and the Australian Bureau of Statistics build authority through independence and the consistent production of credible information. Synthetic media increases the importance of this role, particularly during elections, when uncertainty about official information may weaken confidence in otherwise legitimate processes.

Maintaining this role requires stable funding, institutional independence and the capacity to verify and communicate information clearly. However, support for trusted institutions is only one part of the response. Legal frameworks, technical measures, education, independent media and other civic initiatives each address different aspects of the harm caused by synthetic media. A layered approach is therefore needed to reduce reliance on any single institution or measure while protecting human rights and democratic participation.

In this framework, prevent, respond and remedy describe what the response must do, while legal, technical and civic measures describe the tools and institutions that contribute to those functions.

Finding

Australia requires a coordinated and layered national response that can prevent harm where possible, respond effectively when synthetic media circulates and provide remedies and accountability after harm occurs. No single measure is sufficient. These functions require complementary legal, technical and civic measures, including proportionate laws, technical transparency, digital and civics education, independent media, responsible content creators and trusted public institutions. Measures may contribute to more than one function and should reinforce one another while protecting human rights and democratic participation.

5. Context

This chapter sets out the context for the report. It outlines the research approach and consultation process that informed the analysis.

5.1 Method

This report is based on desk-based research and significant expert consultation. The research component reviewed academic literature, grey literature, policy analysis and Australian legal frameworks relevant to the creation and use of synthetic content. This included consideration of electoral law, online safety regulation, privacy protections and defamation law, alongside international human rights standards relevant to democratic participation.

The consultation component was designed to test and refine the report's analytical framing, rather than to survey public attitudes. Discussions focused on systemic risks, institutional constraints and areas where participants raised similar concerns. Consultation material was analysed thematically to identify patterns relevant to the report's scope.

The report deliberately moves beyond individual incidents of misuse in order to assess structural effects on democratic participation. This approach reflects consultation feedback that case-based responses do not adequately capture the broader democratic implications of synthetic media.

5.2 Consultation

The Commission consulted stakeholders with expertise in technology governance, human rights, national security and electoral administration. Consultations were conducted through meetings and written exchanges during 2026.

Participants included industry, academic experts, civil society organisations, legal bodies and Australian Government agencies with regulatory or security responsibilities. The Commission also engaged with most major social media platforms whose products and governance practices shape how synthetic media is produced and circulated in Australia.

Public sector consultations included bodies responsible for electoral administration, online safety and privacy oversight. Civil society participants contributed expertise in digital rights, gender-based harm and democratic accountability. Academic contributors focused on policy design and information integrity. Platform consultations addressed content governance practices, detection tools and election-period processes.

Consultations were used to understand how different institutions encounter deepfake-related harms and where existing frameworks have recognised limitations. Stakeholders were not asked to reach agreement, and the report does not attribute specific views to individual organisations unless stated elsewhere.

The consultation material informs the report's analysis by clarifying shared constraints and areas of tension that shape Australia's capacity to respond to deepfakes in a way that supports democratic participation and human rights.

6. Appendix

Appendix A: Definitions

Term	Definition
Algorithmic amplification	The increased visibility or reach of content resulting from automated systems that rank, recommend or distribute it.
Content provenance	Information about the origin and history of digital content, including how it was created and subsequently altered. Provenance can assist assessments of authenticity but does not, by itself, establish that a claim or depicted event is true.
Deepfake	Synthetic media which is an AI-generated or altered photo, video or audio of a real person doing or saying something that they did not actually do or say.
Disinformation	Information that is false or misleading and is shared with intent to deceive.
Echo chamber	An information environment in which people predominantly encounter views that reinforce their existing beliefs, with limited exposure to competing perspectives or critical challenge.
Generative artificial intelligence	AI systems that generate new content, including text, images, video or audio, based on patterns learned from training data.
Labelling	The identification of content that has been generated or altered using AI, including through visible markers or embedded data.
Metadata	Information embedded within a digital file that describes its origin or modification, including whether it has been synthetically generated.
Misinformation	Information that is false or misleading but is shared without intent to deceive.
Recommender system	An automated system that selects, ranks or suggests content to users, influencing what they encounter online. Recommendations may draw on users' behaviour, preferences and other signals.
Synthetic Media	An umbrella term for any AI-generated or altered photo, video or audio (including deepfakes).
Watermarking	A technique that embeds an identifying signal in digital content to indicate its origin or AI-generated nature. Watermarks may be visible or detectable using specialised tools and are distinct from information stored only in a file's metadata.

Endnotes

1 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 28.

2 Daniel Innerarity, *Artificial Intelligence and Democracy* (Report, UNESCO, 16 May 2024) 12.

3 Human Rights Council Advisory Committee, *Impact of Disinformation on the Enjoyment and Realization of Human Rights*, UN Doc A/HRC/61/68 (2026) [1].

4 Carnegie Endowment for International Peace, *AI and democracy: Mapping the intersections* (Report, 2026) 3.

5 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 28.

6 Office of the Australian eSafety Commissioner ('eSafety'), *Deepfake trends and challenges* (Position Statement).

7 Lowy Institute, *Lowy institute Poll* (Poll, 2026).

8 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 4 citing Australian Public Service Commission, *Trust and Satisfaction in Australian Democracy: 2023 National Survey* (Survey, 14 March 2024).

9 Centre for Emerging Technology and Security, *AI-Enabled Influence Operations: The Threat to the UK General Election* (Report, May 2024) 10.

10 UNESCO, *Deepfakes and the crisis of knowing* (Webpage) <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing>; See also, ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11-12.

11 See generally eSafety Commissioner, *Recommender Systems and Algorithms – Full Position Statement* (Position Statement, 11 October 2023) 4-6.

12 Cornelia Sindermann, Andrew Cooper & Christian Montag, 'A Short Review on Susceptibility to Falling for Fake Political News' (2020) 36 *Current Opinion in Psychology* 44, 46–47; See also Gillian Murphy et al., 'False Memories for Fake News During Ireland's Abortion Referendum' (2019) 30(10) *Psychological Science* 1449.

13 Gordon Pennycook & David Rand, 'The Psychology of Fake News' (2021) 25(5) *Trends in Cognitive Sciences* 388, 399.

14 UNESCO, *Elections in Digital Times: A Guide for Electoral Practitioners* (Report, 2022) 42.

15 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence* (Interim Report, Commonwealth of Australia, October 2024) 10 [2.17]; See e.g. Ahmet Görgen and Can Saygıner, 'Deepfake Technology, Media, and National Security: The Case of the German Chancellor's Deepfake Video' (2025) 21(51) *Güvenlik Stratejileri Dergisi* 279, 285.

16 Taberez Ahmed Neyazi, et al,. 'Does Trust in Government Moderate the Perception towards Deepfakes? Comparative Perspectives from Asia on the Risks of AI and Misinformation for Democracy' (2024) 41 *Government Information Quarterly* 1, 3.

17 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 12-13.

18 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 12-13.

19 United Nations Human Rights Council, *Impact of Disinformation on the Enjoyment and Realization of Human Rights* (Report, UN Doc A/HRC/61/68, 2026) 5 [16].

20 Gina Romero, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association on Impact of the 2023–2025 "super election" cycle on the rights to freedom of peaceful assembly and of association,* UN Doc A/HRC/59/44 (2025) [1].

21 Gina Romero, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association on Impact of the 2023–2025 "super election" cycle on the rights to freedom of peaceful assembly and of association,* UN Doc A/HRC/59/44 (2025) [1]; V-Dem Institute, *Democracy Report 2026, Unravelling the Democratic Era?* (Report, March 2026) 4; see also Freedom House, *Freedom in the World 2022, The Global Expansion of Authoritarian Rule* (Report, February 2022) 1-16.

22 Bastian Herre *The world has recently become less democratic* (Our World in Data, 6 September 2022, updated April 2026).

23 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 29.

24 United Nations ('UN'), *Democracy and Human Rights: The Role of the UN* (Discussion Paper, 2013) 6.

25 United Nations, Democracy and Human Rights: The Role of the UN (Discussion Paper, 2013) 6; Democracy without Borders, A United Nations Special Rapporteur on Democracy (Report, 2023) 1.

26 *Universal Declaration of Human Rights, GA Res 217A (III), UN GAOR, 3rd sess, 183rd plen mtg, UN Doc A/810 (10 December 1948) art 21; International Covenant on Civil and Political Rights, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) art 25.*

27 *ICCPR* art 19.

28 *ICCPR* art 18.

29 *ICCPR* art 17.

30 *Universal Declaration of Human Rights*, GA Res 217A (III), UN GAOR, 3rd sess, 183rd plen mtg, UN Doc A/810 (10 December 1948) arts 1, 2, 7; *ICCPR* arts 2, 3, 26.

31 *ICCPR* art 19(2).

32 *ICCPR* arts 21-22.

33 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 21.

34 See e.g. Dan Clark and Anna Nicolaou, 'Inside Donald Trump's AI 'slopaganda' machine', *Australian Financial Review* (online, 21 May 2026) <https://www.afr.com/world/north-america/inside-donald-trump-s-ai-slopaganda-machine-20260521-p5zzbh>; Bernd Debusmann Jr, 'Trump deletes post depicting him as Jesus-like figure after backlash', *BBC* (online, 14 Paril 2026) <https://www.bbc.com/news/articles/c17v8y0z9z2o>; Lucia Stein, 'Meme warfare and 'slopaganda' are defining the AI-era Iran war' (online, 22 April 2026) <https://www.abc.net.au/news/2026-04-22/us-israel-strikes-on-iran-propaganda/106585244 https://theconversation.com/slopaganda-wars-how-and-why-the-us-and-iran-are-flooding-the-zone-with-viral-ai-generated-noise-280024>.

35 Human Rights Committee, General Comment No 25: The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service (Art 25), 57th sess, UN Doc CCPR/C/21/Rev.1/Add.7 (12 July 1996) [8], [11].

36 See *International Covenant on Civil and Political Rights,* opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) ('ICCPR') arts 19 & 25.

37 Access Now, *Human Rights in the Age of Artificial Intelligence* (Report, 2018) 25.

38 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 4; See generally Irene Khan, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on Gendered Disinformation and Its Implications for the Right to Freedom of Expression*, UN Doc A/78/288 (2023) [116].

39 See generally, Julie Posetti, et al., *Tipping Point, The Chilling Escalation of Online Violence Against Women in the Public Sphere* (Report, Ending Violence against Women Section, Policy Programme and Intergovernmental Division, UN Women, New York, in the framework of the "ACT to End Violence against Women" Programme, December 2025). *Convention on the Elimination of All Forms of Discrimination against Women*, art 7; Committee on the Elimination of Discrimination against Women, *General Recommendation No 23: Political and Public Life*, UN Doc A/52/38 (1997) [18] (identifying cultural patterns, social attitudes and gender stereotypes among barriers to women's equal participation in political and public life).

40 Irene Khan, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on Gendered Disinformation and Its Implications for the Right to Freedom of Expression*, UN Doc A/78/288 (2023) [50] & [116]; See generally Data & Society Institute, *Online Harassment, Digital Abuse, and Cyberstalking in America* (Report, 2016); See also Amnesty International, *Toxic Twitter – a toxic place for women* (Report, 2018). *Convention on the Elimination of All Forms of Discrimination against Women*, art 7; Committee on the Elimination of Discrimination against Women, *General Recommendation No 23: Political and Public Life*, UN Doc A/52/38 (1997) [18] (identifying cultural patterns, social attitudes and gender stereotypes among barriers to women's equal participation in political and public life).

41 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 15.

42 Irene Khan, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression report on Gendered Disinformation and Its Implications for the Right to Freedom of Expression*, UN Doc A/78/288 (2023) [45].

43 National Democratic Institute for International Affairs, *Tweets that chill: Analysing Online Violence Against Women in Politics* (Report, 2019) 21-25.

44 Plan International, *The truth gap: how misinformation and disinformation online affect the lives, learning and leadership of girls and young women* (Report, 2021) 6.

45 Asher Flynn, et al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non Consensually Created and Shared Sexualized Deepfake Imagery' (2025) 00(0) *Journal of Interpersonal Violence* 1, 12.

46 Asher Flynn et al, 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of NonConsensually Created and Shared Sexualized Deepfake Imagery' (2025) *Journal of Interpersonal Violence*.

47 UNESCO, *Elections in Digital Times: A Guide for Electoral Practitioners* (Report, 2022) 46; see also Asher Flynn, et al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non Consensually Created and Shared Sexualized Deepfake Imagery' (2025) 00(0) *Journal of Interpersonal Violence* 1, 6.

48 Europol, *Malicious Uses and Abuses of Artificial Intelligence* (Trend Micro Research, 2020) 52 ; see also Asher Flynn, et al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non Consensually Created and Shared Sexualized Deepfake Imagery' (2025) 00(0) *Journal of Interpersonal Violence* 1, 1.

49 Noelle Martin, 'Online safety regulation of deepfake abuse: a case study on Australia's eSafety Commissioner' (2025) 34(1) *Griffith Law Review* 23, 23.

50 See generally Deeptrace Labs, *The State of Deepfakes: Landscape, Threats, and Impact* (Report, 2019).

51 Julie Inman Grant, 'Addressing deepfake image-based abuse', *eSafety* (Online, 24 July 2024) <https://www.esafety.gov.au/newsroom/blogs/addressing-deepfake-image-based-abuse>.

52 Nadeem Badshah, 'Nearly 4,000 celebrities found to be victims of deepfake pornography', *The Guardian* (online, 22 March 2024) <https://www.theguardian.com/technology/2024/mar/21/celebrities-victims-of-deepfake-pornography>.

53 Churchill Fellowship, *Cara Hunter* (Webpage) <https://www.churchillfellowship.org/ideas-experts/fellows-directory/cara-hunter/>.

54 Anna Moore, ''It was extremely pornographic': Cara Hunter on the deepfake video that nearly ended her political career', *The Guardian* (online, 01 December 2025) <https://www.theguardian.com/society/ng-interactive/2025/dec/01/it-was-extremely-pornographic-cara-hunter-on-the-deepfake-video-that-nearly-ended-her-political-career>.

55 Anna Moore, ''It was extremely pornographic': Cara Hunter on the deepfake video that nearly ended her political career', *The Guardian* (online, 01 December 2025) <https://www.theguardian.com/society/ng-interactive/2025/dec/01/it-was-extremely-pornographic-cara-hunter-on-the-deepfake-video-that-nearly-ended-her-political-career>.

56 Anna Moore, ''It was extremely pornographic': Cara Hunter on the deepfake video that nearly ended her political career', *The Guardian* (online, 01 December 2025) <https://www.theguardian.com/society/ng-interactive/2025/dec/01/it-was-extremely-pornographic-cara-hunter-on-the-deepfake-video-that-nearly-ended-her-political-career>.

57 Churchill Fellowship, *Cara Hunter* (Webpage) <https://www.churchillfellowship.org/ideas-experts/fellows-directory/cara-hunter/>.

58 Asher Flynn, et al., 'Sexualized Deepfake Abuse: Perpetrator and Victim Perspectives on the Motivations and Forms of Non Consensually Created and Shared Sexualized Deepfake Imagery' (2025) 00(0) *Journal of Interpersonal Violence* 9042, 9043.

59 Jason Wilson, 'Hundreds of nonconsensual AI images being created by Grok on X, data shows', *The Guardian* (Online, 09 January 2026) <https://www.theguardian.com/technology/2026/jan/08/grok-x-nonconsensual-images>.

60 Agence France-Presse, 'ChatGPT to generate adult erotica from December as OpenAI eases content restrictions', *The Guardian* (online, 15 October 2025) https://www.theguardian.com/technology/2025/oct/14/openai-chatgpt-adult-erotic-content; Reuters, 'OpenAI indefinitely pauses plans to release erotic chatbot, FT says', *Reuters* (online, 26 March 2026) <https://www.reuters.com/business/openai-indefinitely-pauses-plans-release-erotic-chatbot-ft-says-2026-03-26/>.

61 *eSafety Commissioner v Rotondo (No 4)* [2025] FCA 1191 [12].

62 *eSafety Commissioner v Rotondo (No 4)* [2025] FCA 1191 [12].

63 See generally Committee on the Elimination of Discrimination against Women, *General Recommendation No. 19: Violence against Women* (1992, UN Doc A/47/38); Committee on the Elimination of Discrimination against Women, *General Recommendation No. 35 on GenderBased Violence against Women, Updating General Recommendation No. 19* (2017, UN Doc CEDAW/C/GC/35).

64 UNHRC, *General Comment No. 25: Article 25 (Participation in Public Affairs and the Right to Vote), The Right to Participate in Public Affairs, Voting Rights and the Right of Equal Access to Public Service* (12 July 1996, UN Doc CCPR/C/21/Rev.1/Add.7) 1 [3].

65 Europol, *Malicious Uses and Abuses of Artificial Intelligence* (Report, 2020) 52.

66 Ashwini K P, *Contemporary forms of racism, racial discrimination, xenophobia and related intolerance* (Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, 2024, UN Doc A/HRC/56/68) 7 [20].

67 Irene Khan, *Gendered Disinformation and Its Implications for the Right to Freedom of Expression* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2023, UN Doc A/78/288) 11 [52] & 6 [17].

68 Irene Khan, *Gendered Disinformation and Its Implications for the Right to Freedom of Expression* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2023, UN Doc A/78/288) 11 [52].

69 Irene Khan, *Disinformation and freedom of opinion and expression during armed conflicts* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2022, UN Doc A/77/288) 6 [18].

70 Human Rights Committee, *General Comment No 36: Article 6 (Right to Life)* (2018, UN Doc CCPR/C/GC/36) 1 [3].

71 Human Rights Committee, *General Comment No 36: Article 6 (Right to Life)* (2018, UN Doc CCPR/C/GC/36) 1 [2].

72 Human Rights Committee, *General Comment No 36: Article 6 (Right to Life)* (2018, UN Doc CCPR/C/GC/36) 5 [23] which expands the right to include LGBTQI+ and trans and gender-diverse people.

73 House Select Committee on Social Media and Online Safety, *Social Media and Online Safety* (Report, 2022) 29-44.

74 Human Rights Committee, *General Comment No 35: Article 9 (Liberty and Security of Person)* (2014, UN Doc CCPR/C/GC/35) 16 [55].

75 UNESCO, Submission to the UN Human Rights Council inquiry on the *Impact of disinformation on the enjoyment and realization of human rights* (Submission, 2025) 2; House Select Committee on Social Media and Online Safety, *Social Media and Online Safety* (Report, 2022) 29-44.

76 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 30.

77 UNESCO, *Elections in Digital Times: A Guide for Electoral Practitioners* (Report, 2022) 34.

78 AI Incident Database, *Incident 1336: BJP Used Deepfake Videos of Manoj Tiwari to Target Haryanvi Voters in 2020 Delhi Election* (Webpage) <https://incidentdatabase.ai/cite/1336>.

79 Sukant Khurana, 'Deepfakes: Fracturing India's Social Fabric', *Medium* (online, 19 June 2024) <https://medium.com/@sukantkhurana/deepfakes-fracturing-indias-social-fabric-9493d6c37112>.

80 Sukant Khurana, 'Deepfakes: Fracturing India's Social Fabric', *Medium* (online, 19 June 2024) <https://medium.com/@sukantkhurana/deepfakes-fracturing-indias-social-fabric-9493d6c37112>.

81 Sukant Khurana, 'Deepfakes: Fracturing India's Social Fabric', *Medium* (online, 19 June 2024) <https://medium.com/@sukantkhurana/deepfakes-fracturing-indias-social-fabric-9493d6c37112>.

82 Amnesty International, Thr*ee out five young activists face online harassment globally for posting human rights content* (webpage) <https://www.amnesty.org/en/latest/news/2024/07/three-out-five-young-activists-face-online-harassment-globally-for-posting-human-rights-content/>; See generally, Julie Posetti, et al., *Tipping Point, The Chilling Escalation of Online Violence Against Women in the Public Sphere* (Report, Ending Violence against Women Section, Policy Programme and Intergovernmental Division, UN Women, New York, in the framework of the "ACT to End Violence against Women" Programme, December 2025); See also Amnesty International, *Online Violence* (webpage) <https://www.amnesty.org/en/what-we-do/technology/online-violence/>.

83 Amnesty International, Thr*ee out five young activists face online harassment globally for posting human rights content* (webpage) <https://www.amnesty.org/en/latest/news/2024/07/three-out-five-young-activists-face-online-harassment-globally-for-posting-human-rights-content/>.

84 Human Rights Committee, *General Comment No 34: Article 19 (Freedoms of Opinion and Expression)* (2011, UN Doc CCPR/C/GC/34) 1 [2].

85 See generally, Julie Posetti, et al., *Tipping Point, The Chilling Escalation of Online Violence Against Women in the Public Sphere* (Report, Ending Violence against Women Section, Policy Programme and Intergovernmental Division, UN Women, New York, in the framework of the "ACT to End Violence against Women" Programme, December 2025); See also Amnesty International, *Online Violence* (webpage) <https://www.amnesty.org/en/what-we-do/technology/online-violence/>.

86 UNESCO, Submission to the UN Human Rights Council inquiry on the *Impact of disinformation on the enjoyment and realization of human rights* (Submission, 2025) 2.

87 Synthetic media can be a form of free expression as *ICCPR* art 19 covers 'all forms of audio-visual as well as electronic and internet-based modes of expression'; Human Rights Committee, *General Comment No 34: Article 19 (Freedoms of Opinion and Expression)* (2011, UN Doc CCPR/C/GC/34) 3 [12].

88 *ICCPR* art 19 covers the freedom to '... impart information and ideas of all kinds ... in the form of art'. While there is much debate as to whether synthetic media is 'art' in the traditional sense is contested - satire, parody and caricature are generally classified as artistic, cultural and political commentary, they are sheltered under the umbrella of 'ideas of all kinds' delivered 'in the form of art'; see further Human Rights Committee, *General Comment No 34: Article 19 (Freedoms of Opinion and Expression)* (2011, UN Doc CCPR/C/GC/34) 3 [11]; see also Alex Barber, 'Freedom of Expression Meets Deepfakes' (2023) 202(40) *Synthese* 1, 3.

89 Michael Stott & Jyotsna Singh, 'Viral 'cockroach party' crawls under Narendra Modi's skin', *Australia Financial Review* (online, 26 May 2026) <https://www.afr.com/world/asia/viral-cockroach-party-crawls-under-narendra-modi-s-skin-20260526-p600tm>; Zoya Mateen, 'India has a new political superstar - a cockroach', *BBC* (online, 21 May 2026) <https://www.bbc.com/news/articles/cz72y11jjq1o>; Hannah Ellis-Petersen, ''What if all cockroaches came together?' The youth movement threatening to shake up India's politics', *The Guardian* (online, 08 June 2026) <https://www.theguardian.com/world/2026/jun/08/cockroach-janta-party-youth-movement-india-politics>.

90 Michael Stott & Jyotsna Singh, 'Viral 'cockroach party' crawls under Narendra Modi's skin', *Australia Financial Review* (online, 26 May 2026) <https://www.afr.com/world/asia/viral-cockroach-party-crawls-under-narendra-modi-s-skin-20260526-p600tm>; Zoya Mateen, 'India has a new political superstar - a cockroach', *BBC* (online, 21 May 2026) <https://www.bbc.com/news/articles/cz72y11jjq1o>; Hannah Ellis-Petersen, ''What if all cockroaches came together?' The youth movement threatening to shake up India's politics', *The Guardian* (online, 08 June 2026) <https://www.theguardian.com/world/2026/jun/08/cockroach-janta-party-youth-movement-india-politics>.

91 Michael Stott & Jyotsna Singh, 'Viral 'cockroach party' crawls under Narendra Modi's skin', *Australia Financial Review* (online, 26 May 2026) <https://www.afr.com/world/asia/viral-cockroach-party-crawls-under-narendra-modi-s-skin-20260526-p600tm>; Zoya Mateen, 'India has a new political superstar - a cockroach', *BBC* (online, 21 May 2026) <https://www.bbc.com/news/articles/cz72y11jjq1o>; Hannah Ellis-Petersen, ''What if all cockroaches came together?' The youth movement threatening to shake up India's politics', *The Guardian* (online, 08 June 2026) <https://www.theguardian.com/world/2026/jun/08/cockroach-janta-party-youth-movement-india-politics>.

92 Connor Greene, What Is the Youth-Led 'Cockroach Movement' Staging Mass Protests in India?, *Time* (Online, 25 July 2026) <https://time.com/article/2026/07/22/india-cockroach-janta-party-protest-movement/>.

93 See e.g. Cockroach Janta Party, *Cockroach Janta Party* (webpage) <https://www.cockroachjantaparty.org/>.

94 Lindsey Wilkerson, 'Still Waters Run Deep(fakes): The Rising Concerns of "Deepfake" Technology and Its Influence on Democracy and the First Amendment' (2021) 86(1) *Missouri Law Review* 1, 7.

95 See e.g. Lindsey Wilkerson, 'Still Waters Run Deep(fakes): The Rising Concerns of "Deepfake" Technology and Its Influence on Democracy and the First Amendment' (2021) 86(1) *Missouri Law Review* 1, 20.

96 Tess Bennett, 'Albanese mocked by entrepreneurs using AI images to fight CGT changes', *The Australian Financial Review* (online, 18 May 2026) <https://www.afr.com/technology/albanese-mocked-by-entrepreneurs-using-ai-images-to-fight-cgt-changes-20260518-p5zy6a>.

97 David Adams, 'Albanese thanks business owners for "very nice photos" in CGT memes', *Smart Company* (Online, 19 May 2026) <https://www.smartcompany.com.au/federal-budget-2026/albanese-thanks-business-owners-for-very-nice-photos-in-cgt-memes/>.

98 Dan Clark and Anna Nicolaou, 'Inside Donald Trump's AI 'slopaganda' machine', *Australian Financial Review* (online, 21 May 2026) <https://www.afr.com/world/north-america/inside-donald-trump-s-ai-slopaganda-machine-20260521-p5zzbh>; Bernd Debusmann Jr, 'Trump deletes post depicting him as Jesus-like figure after backlash', *BBC* (online, 14 Paril 2026) <https://www.bbc.com/news/articles/c17v8y0z9z2o>; Lucia Stein, 'Meme warfare and 'slopaganda' are defining the AI-era Iran war' (online, 22 April 2026) <https://www.abc.net.au/news/2026-04-22/us-israel-strikes-on-iran-propaganda/106585244 https://theconversation.com/slopaganda-wars-how-and-why-the-us-and-iran-are-flooding-the-zone-with-viral-ai-generated-noise-280024>.

99 David Kaye, *Surveillance and human rights* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2019, UN Doc A/HRC/41/35) 4 [11]; see also Maina Kiai, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2012, UN Doc A/HRC/20/27); Clément Nyaletsossi Voule, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2018, UN Doc A/HRC/38/34).

100 See generally Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 15

101 Helena Horton, et al., 'Politicians try to calm tensions inflamed by social media after stabbing in Belfast', *The Guardian* (online, 10 June 2026) <https://www.theguardian.com/uk-news/2026/jun/09/politicians-tensions-musk-social-media-stabbing-belfast>; ABC, 'Keir Starmer urges calm after protests over handcuffed Henry Nowak's death', *ABC* (online, 04 June 2026) <https://www.abc.net.au/news/2026-06-04/starmer-urges-calm-after-protests-over-student-henry-nowak-death/106757326>.

102 Jason Okundaye, 'The right has created a false reality – fuelled by toxic images delivered straight to your phone', *The Guardian* (online, 13 June 2026) <https://www.theguardian.com/commentisfree/2026/jun/13/right-false-reality-toxic-images-riots-belfast-southampton>.

103 The London School of Economics and Political Science, *How fake AI images sparked the Southport riots* (webpage) <https://www.lse.ac.uk/lse-player/how-fake-ai-images-sparked-the-southport-riots>.

104 AAP, *Fake video showing massive March for Australia crowds was created with AI* (webpage) <https://aapnews.aap.com.au/news/fake-video-showing-massive-march-for-australia-crowds-was-created-with-ai>.

105 Matthew Elmas, 'Palestine protest scuffle image an AI fabrication', *AAP FactCheck* (online, 26 March 2026) <https://www.aap.com.au/factcheck/palestine-protest-scuffle-image-an-ai-fabrication/>.

106 *ICCPR* art 18; *UDHR* art 18.

107 Ahmed Shaheed, *Freedom of Thought* (Report of the Special Rapporteur on freedom of religion or belief, 2021, UN Doc A/76/380) 3 [4].

108 Ahmed Shaheed, *Freedom of Thought* (Report of the Special Rapporteur on freedom of religion or belief, 2021, UN Doc A/76/380) 3 [5].

109 Ahmed Shaheed, *Freedom of Thought* (Report of the Special Rapporteur on freedom of religion or belief, 2021, UN Doc A/76/380) 11 [35]. The line appears to be when persuasion moves into the realm of manipulation that it hits the threshold of impermissible alteration.

110 Ahmed Shaheed, *Freedom of Thought* (Report of the Special Rapporteur on freedom of religion or belief, 2021, UN Doc A/76/380) 11.

111 Ahmed Shaheed, *Freedom of Thought* (Report of the Special Rapporteur on freedom of religion or belief, 2021, UN Doc A/76/380) 10-11 [29]-[39].

112 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 3.

113 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 3.

114 UN Department of Global Communications, *From Principles to Practice: Strengthening Information Integrity (Issue Brief 1)* (Report, September 2025) 3.

115 UN Department of Global Communications, *From Principles to Practice: Strengthening Information Integrity (Issue Brief 1)* (Report, September 2025) 3.

116 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 3.

117 See generally Carnegie Endowment for International Peace, *AI and Democracy: Mapping the Intersections* (Report, January 2026).

118 See generally UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024).

119 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 3.

120 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 8-9.

121 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 10-11.

122 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 12-13.

123 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 14-15.

124 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 16-17.

125 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 36-39.

126 Nicholas Biddle & Matthew Gray, *Australian Views Towards Democracy: Comparisons through time and with the rest of the region* (Report, 2020) 4.

127 Nicholas Biddle & Matthew Gray, *Australian Views Towards Democracy: Comparisons through time and with the rest of the region* (Report, 2020) 3-4; See generally Sarah Cameron and Ian McAllister, *Trends in Australian Political Opinion: Results from the Australian Election Study 1987–2022* (Report, December 2022); See also Ruth Dassonneville and Ian McAllister, 'Explaining the Decline of Political Trust in Australia' (2021) 56(2) *Australian Journal of Political Science* 280.

128 Sarah Cameron, et al., *The 2022 Australian Federal Election: Results from the Australian Election Study* (Report, December 2022) 27-28.

129 Irene Khan, *Disinformation and freedom of opinion and expression during armed conflicts* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2022, UN Doc A/77/288) 9 [26].

130 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence* (Final Report, Commonwealth of Australia, November 2024) 19 [2.34].

131 UNESCO, Submission to the UN Human Rights Council inquiry on the *Impact of disinformation on the enjoyment and realization of human rights* (Submission, 2025) 4.

132 Aditya Kalra, Munsif Vengattil & Dhwani Pandya, 'Deepfakes of Bollywood stars spark worries of AI meddling in India election', *Reuters* (Online, 22 April 2024) <https://www.reuters.com/world/india/deepfakes-bollywood-stars-spark-worries-ai-meddling-india-election-2024-04-22/>.

133 Aditya Kalra, Munsif Vengattil & Dhwani Pandya, 'Deepfakes of Bollywood stars spark worries of AI meddling in India election', *Reuters* (Online, 22 April 2024) <https://www.reuters.com/world/india/deepfakes-bollywood-stars-spark-worries-ai-meddling-india-election-2024-04-22/>.

134 Aditya Kalra, Munsif Vengattil & Dhwani Pandya, 'Deepfakes of Bollywood stars spark worries of AI meddling in India election', *Reuters* (Online, 22 April 2024) <https://www.reuters.com/world/india/deepfakes-bollywood-stars-spark-worries-ai-meddling-india-election-2024-04-22/>.

135 Aditya Kalra, Munsif Vengattil & Dhwani Pandya, 'Deepfakes of Bollywood stars spark worries of AI meddling in India election', *Reuters* (Online, 22 April 2024) <https://www.reuters.com/world/india/deepfakes-bollywood-stars-spark-worries-ai-meddling-india-election-2024-04-22/>.

136 Ukrinform, 'MFA Ukraine presents "digital spokesperson" for consular affairs', *Ukrinform* (online, 1 May 2024) <https://www.ukrinform.net/rubric-society/3858754-mfa-ukraine-presents-digital-spokesperson-for-consular-affairs.html>; Leslie Katz, 'Ukraine Government Appoints AI-Generated Spokesperson', *Forbes* (online, 3 May 2024) <https://www.forbes.com/sites/lesliekatz/2024/05/03/ukraine-foreign-ministry-introduces-ai-generated-spokesperson/>.

137 Matthew Elmas, 'AI used to put words into the mouth of police chief following Bondi attack', *AAP* (Online, 23 December 2025) <https://www.aap.com.au/factcheck/ai-used-to-put-words-into-the-mouth-of-police-chief-following-bondi-attack/>.

138 Australian Federal Police, 'Clip of Australian police announcing arrest of four Indian nationals over Bondi attack is AI-generated' (Media Release, 01 January 2026).

139 See e.g. Ahmet Görgen and Can Saygıner, 'Deepfake Technology, Media, and National Security: The Case of the German Chancellor's Deepfake Video' (2025) 21(51) *Güvenlik Stratejileri Dergisi* 279, 280-283.

140 Sanela Veljković, Milica Ćurčić & Ilija Gavrilović, 'Dark Sides of Deepfake Technology' (2024) 72(3) *Vojnotehnički Glasnik* 1441, 1449.

141 Art Jipson, 'Charlie Kirk and the making of an AI-generated martyr', *The Conversation* (Online, 26 September 2025) <https://theconversation.com/charlie-kirk-and-the-making-of-an-ai-generated-martyr-265834>.

142 See e.g. Steve Rose, 'The Slopaganda Era: 10 AI Images Posted by the White House – and What They Teach Us', *The Guardian* (online, 29 January 2026) <https://www.theguardian.com/us-news/2026/jan/29/the-slopaganda-era-10-ai-images-posted-by-the-white-house-and-what-they-teach-us>.

143 Dan Clark and Anna Nicolaou, 'Inside Donald Trump's AI 'slopaganda' machine', *Australian Financial Review* (online, 21 May 2026) <https://www.afr.com/world/north-america/inside-donald-trump-s-ai-slopaganda-machine-20260521-p5zzbh>; Bernd Debusmann Jr, 'Trump deletes post depicting him as Jesus-like figure after backlash', *BBC* (online, 14 Paril 2026) <https://www.bbc.com/news/articles/c17v8y0z9z2o>.

144 Joe Sommerlad, 'Trump posts AI video of Obama being arrested after Gabbard's coup claims: 'No one is above the law'', *Independent* (Online, 22 July 2025) <https://www.independent.co.uk/news/world/americas/us-politics/trump-obama-arrest-ai-video-truth-social-b2792710.html>.

145 Joe Sommerlad, 'Trump posts AI video of Obama being arrested after Gabbard's coup claims: 'No one is above the law'', *Independent* (Online, 22 July 2025) <https://www.independent.co.uk/news/world/americas/us-politics/trump-obama-arrest-ai-video-truth-social-b2792710.html>.

146 Australian Competition and Consumer Commission, 'ACCC takes action over alleged misleading conduct by Meta for publishing scam celebrity crypto ads on Facebook' (media release, 18 March 2022).

147 ScamWatch, 'Scam alert: Fake celebrity online investment scams' (Media Release, 01 March 2024).

148 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence* (Interim Report, Commonwealth of Australia, October 2024) 11 [2.23] quoting ABC, Submission 117, 2; International Institute for Democracy and Electoral Assistance, *Artificial Intelligence for Electoral Management* (Report, 2024) 43; Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic AudioVisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1 *Digital Society* 19.

149 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic AudioVisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1 *Digital Society* 19; Igor Derysh, 'GOP House candidate publishes lengthy report claiming George Floyd's killing was a "deepfake" hoax', *Salon* (Online, 24 June 2020) <https://www.salon.com/2020/06/24/gop-house-candidate-publishes-lengthy-report-claiming-george-floyds-killing-was-a-deepfake-hoax/>.

150 Dr. Winnie Heartstrong for Congress, *X Post* (Online, 15 June 2020) <https://x.com/realDrWinnie/status/1272526523858378753?refsrc=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1272526523858378753%7Ctwgr%5Ee521e86fb4eecff84f9125531fcb5e5e040abd2a%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.salon.com%2F2020%2F06%2F24%2Fgop-house-candidate-publishes-lengthy-report-claiming-george-floyds-killing-was-a-deepfake-hoax%2F>; Igor Derysh, 'GOP House candidate publishes lengthy report claiming George Floyd's killing was a "deepfake" hoax', *Salon* (Online, 24 June 2020) <https://www.salon.com/2020/06/24/gop-house-candidate-publishes-lengthy-report-claiming-george-floyds-killing-was-a-deepfake-hoax/>.

151 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic AudioVisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1 *Digital Society* 19-20.

152 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic AudioVisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1 *Digital Society* 20.

153 The New York Times, Netanyahu Posts 'Proof of Life' Video as AI Sows Doubts about what's Real', *The New York Times* (Online, 17 March 2026) <https://www.nytimes.com/2026/03/17/technology/netanyahu-ai-video-iran-israel.html>.

154 The New York Times, Netanyahu Posts 'Proof of Life' Video as AI Sows Doubts about what's Real', *The New York Times* (Online, 17 March 2026) <https://www.nytimes.com/2026/03/17/technology/netanyahu-ai-video-iran-israel.html>.

155 UN, *United Nations Global Principles for Information Integrity: Recommendations for Multistakeholder Action* (Report, June 2024) 14-15.

156 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 24.

157 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 24.

158 RedTechStaff, 'Australian podcast audience tops 4.6 million for news in June', *RedTech* (Online, 28 June 2026) <https://www.redtech.pro/australian-podcast-ranker-june-2026/>.

159 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 39.

160 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 28.

161 Strengthening Democracy Taskforce, *Strengthening Australian democracy: A practical agenda for democratic resilience* (Report, 2024) 39.

162 Irene Khan, *Freedom of Opinion and Expression* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/47/52, 2021) 8 [20].

163 UNESCO, *Behind the Screens: Insights from Digital Content Creators – Understanding Their Intentions, Practices and Challenges* (Report, 2024) 7.

164 University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 11.

165 University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 11.

166 University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 11.

167 Australian Public Service Commission, *Trust and Satisfaction in Australian Democracy Survey Report* (Report, 2023) 12.

168 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 21.

169 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 20-21.

170 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 20-21.

171 Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 20-21.

172 Joe Sommerlad, 'Chris Cuomo trolled for falling for AOC deepfake video about Sydney Sweeney', *The Guardian* (Online, 7 August 2025) <https://www.theguardian.com/us-news/2025/aug/07/chris-cuomo-alexandria-ocasio-cortez-deepfake>.

173 UNESCO, Submission to the UN Human Rights Council inquiry on the *Impact of disinformation on the enjoyment and realization of human rights* (Submission, 2025) 1.

174 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 44-45.

175 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 24.

176 University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 12.

177 University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 11.

178 Lee Fong Yee et al., 'Social Platforms in the Deepfake Age: Navigating Media Trust through Media Literacy' (2025) *International Journal on Informatics Visualization* 176, 176; Europol, *Malicious Uses and Abuses of Artificial Intelligence* (Report, 2024) 53.

179 UNESCO, *Elections in Digital Times: A Guide for Electoral Practitioners* (Report, 2022) 38.

180 Senate Select Committee on Foreign Interference through Social Media, *Foreign Interference through Social Media* (Final Report, Commonwealth of Australia, August 2023) (xi).

181 Dr Jake Wallis, International Cyber Policy Centre, Australian Strategic Policy Institute ('ASPI'), *Committee Hansard* (22 June 2020) 10.

182 ASPI, *Cyber-enabled Foreign Interference in Elections and Referendums* (Report, 2020) 8.

183 Australian Security Intelligence Organisation, *2021-22 Annual Report* (Report, 19 October 2021) 4.

184 Australian Security Intelligence Organisation, *Annual Report 2024–25* (2025) <https://www.transparency.gov.au/publications/home-affairs/australian_security_intelligence_organisation/australian-security-intelligence-organisation-annual-report-2024-25/director-general%E2%80%99s-review>.

185 Australian Security Intelligence Organisation, *2024-25 Annual Report* (Report, 2025).

186 Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 32.

187 Mike Burgess AM, *ASIO Annual Threat Assessment 2025* (Speech, 19 February 2025).

188 ABC News, 'Verify: Disinformation and deepfakes after Bondi attack', *ABC News* (Online, 18 December 2025) <https://www.abc.net.au/news/2025-12-18/verify-disinformation-and-deepfakes-after-bondi-attack/106154250>; Sarah Smaje, et al., 'Bondi attack deepfake Guardian AI viral video', *The Guardian* (Online, 13 January 2026) <https://www.theguardian.com/australia-news/video/2026/jan/13/bondi-attack-deepfake-guardian-ai-viral-video-ntwnfb>.

189 Maryanne Taouk, Michael Workman and Jonathan Hair, 'Racist and antisemitic false information spreads online following Bondi Beach terrorism attack', *ABC News* (Online, 17 December 2025) <https://www.abc.net.au/news/2025-12-17/abc-news-verify-misinformation-bondi-terrorist-attack/106150286>.

190 ABC News, 'Verify: Disinformation and deepfakes after Bondi attack', *ABC News* (Online, 18 December 2025) <https://www.abc.net.au/news/2025-12-18/verify-disinformation-and-deepfakes-after-bondi-attack/106154250>.

191 Phoebe Pin, 'Bondi terror attack survivor tells antisemitism royal commission he received 'influx of hate'', *ABC News* (Online, 29 June 2026) <https://www.abc.net.au/news/2026-06-29/bondi-antisemitism-royal-commission-social-media-hate/106853944>.

192 ABC News, 'Verify: Disinformation and deepfakes after Bondi attack', *ABC News* (Online, 18 December 2025) <https://www.abc.net.au/news/2025-12-18/verify-disinformation-and-deepfakes-after-bondi-attack/106154250>.

193 Select Committee on Foreign Interference through Social Media, *Foreign Interference through Social Media* (First Interim Report, Commonwealth of Australia, December 2021) 32 [4.6].

194 Responsible Technology Australia, Submission No 17 to Senate, *Select Committee on Foreign Interference through Social Media* (2021) 2.

195 Europol, *Malicious Uses and Abuses of Artificial Intelligence* (Report, 2024) 52; See e.g. Ahmet Görgen and Can Saygıner, 'Deepfake Technology, Media, and National Security: The Case of the German Chancellor's Deepfake Video' (2025) 21(51) *Güvenlik Stratejileri Dergisi* 279, 283; Taberez Ahmed Neyazi, et al,. 'Does Trust in Government Moderate the Perception towards Deepfakes? Comparative Perspectives from Asia on the Risks of AI and Misinformation for Democracy' (2024) 41 *Government Information Quarterly* 1, 1.

196 Centre for Emerging Technology and Security, *AI-Enabled Influence Operations: The Threat to the UK General Elections* (Report, May 2024) 10.

197 David Gilbert, 'Russian Propaganda Unit Appears to Be Behind Spread of False Tim Walz Sexual Abuse Claims', *Wired* (online, 21 October 2024) <https://www.wired.com/story/russian-propaganda-unit-storm-1516-false-tim-walz-sexual-abuse-claims/>.

198 The Washington Post, 'Viral attack on Tim Walz features fake former student', *The Washington Post* (online, 21 October 2024) <https://www.washingtonpost.com/investigations/2024/10/21/tim-walz-matthew-metro-video/>; David Gilbert, 'Russian Propaganda Unit Appears to Be Behind Spread of False Tim Walz Sexual Abuse Claims', *Wired* (online, 21 October 2024) <https://www.wired.com/story/russian-propaganda-unit-storm-1516-false-tim-walz-sexual-abuse-claims/>.

199 The Washington Post, 'Viral attack on Tim Walz features fake former student', *The Washington Post* (online, 21 October 2024) <https://www.washingtonpost.com/investigations/2024/10/21/tim-walz-matthew-metro-video/>.

200 The Washington Post, 'Viral attack on Tim Walz features fake former student', *The Washington Post* (online, 21 October 2024) <https://www.washingtonpost.com/investigations/2024/10/21/tim-walz-matthew-metro-video/>.

201 Europol, *Malicious Uses and Abuses of Artificial Intelligence* (Report, 2024) 52 & 60.

202 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence (AI)* (Interim Report, Commonwealth of Australia) 13 [2.32].

203 Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform' (2021) 44(3) *University of New South Wales Law Journal* 983, 987.

204 Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform' (2021) 44(3) *University of New South Wales Law Journal* 983, 987.

205 John Joseph Twomey, Conor Linehan & Gillian Murphy, 'Deepfakes in warfare: new concerns emerge from their use around the Russian invasion of Ukraine', *The Conversation* (27 October 2023) <https://theconversation.com/deepfakes-in-warfare-new-concerns-emerge-from-their-use-around-the-russian-invasion-of-ukraine-216393>; Matthew D Weiner, 'Destined to Deceive: The Need to Regulate Deepfakes with a Foreseeable Harm Standard' (2024) 122 *Michigan Law Review* 771, 775; Maria Pawelec, 'Deepfakes and Democracy (Theory): How Synthetic AudioVisual Media for Disinformation and Hate Speech Threaten Core Democratic Functions' (2022) 1(19) *Digital Society* 1, 17.

206 Matthew Holroyd & Fola Olorunselu, 'Deepfake Zelenskyy surrender video is the 'first intentionally used' in Ukraine war', *Euro News* (Online, 16 March 2022) <https://www.euronews.com/my-europe/2022/03/16/deepfake-zelenskyy-surrender-video-is-the-first-intentionally-used-in-ukraine-war>.

207 National Security: The Case of the German Chancellor's Deepfake Video' (2025) 21(51) *Güvenlik Stratejileri Dergisi* 279, 285.

Taberez Ahmed Neyazi, et al,. 'Does Trust in Government Moderate the Perception towards Deepfakes? Comparative Perspectives from Asia on the Risks of AI and Misinformation for Democracy' (2024) 41 *Government Information Quarterly* 1, 3; See e.g. from Germany Ahmet Görgen and Can Saygıner, 'Deepfake Technology, Media, and National Security: The Case of the German Chancellor's Deepfake Video' (2025) 21(51) *Güvenlik Stratejileri Dergisi* 279, 285.

208 See Federal Bureau of Investigation, *Private Industry Notification* (PIN No 20210310, 10 March 2021).

209 ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11-12.

210 ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11-12.

211 ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11-12.

212 ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11.

213 ASPI, *Weaponised Deep Fakes: National Security and Democracy* (Report, 2020) 11-12.

214 *Commonwealth Electoral Act 1918* (Cth) s 329.

215 The narrow interpretation of *Commonwealth Electoral Act 1918* (Cth) s 328 in *Evans v Crichton-Browne* (1981) 147 CLR 169 was affirmed in *Peebles v Honourable Tony Burke* [2010] FCA 838; *Referendum (Machinery Provisions) Act 1984* (Cth) s 122 uses substantively the same wording. The High Court has held that the wording 'in relation to the casting of a vote' is narrow and applies only in respect of the process of casting a vote (rather than forming a political judgement or influencing an election); See *Abbotto v Australian Electoral Commission* [1997] HCA 18; *Garbett v Liu* (2019) 375 ALR 117.

216 Yee-Fui Ng, *Truth in Political Advertising Laws: Design, Operation, Effectiveness and Recommendations for Reform* (Final Report, December 2024) 5, 23–24.

217 Yee-Fui Ng, *Truth in Political Advertising Laws: Design, Operation, Effectiveness and Recommendations for Reform* (Final Report, December 2024) 23–25.

218 Yee-Fui Ng, *Truth in Political Advertising Laws: Design, Operation, Effectiveness and Recommendations for Reform* (Final Report, December 2024) 48.

219 Yee-Fui Ng, *Truth in Political Advertising Laws: Design, Operation, Effectiveness and Recommendations for Reform* (Final Report, December 2024) 34–37, 62–63.

220 See especially *Electoral Legislation Amendment (Electoral Communications) Bill 2024* (Cth) ss 321M & 321MA.

221 See especially *Electoral Legislation Amendment (Electoral Communications) Bill 2024* (Cth) ss 321M & 321MA.

222 See generally *Electoral Legislation Amendment (Electoral Communications) Bill 2024* (Cth).

223 See *Electoral Act 2004* (Tas) s 197; *Electoral Act 1992* (ACT) s 297A.

224 *Electoral Act 1985* (SA) s 115B.

225 *Electoral Act 1985* (SA) ss 115B(1) & 115C(1)(c).

226 *Electoral Act 1985* (SA) s 115D.

227 Australia institute, *Truth in political advertising* (webpage) <https://australiainstitute.org.au/initiative/truth-in-political-advertising/implementation/>.

228 Seungmin (Helen) Lee , 'AI and Elections: Lessons From South Korea', *The Diplomat* (online, 13 May 2024) <https://thediplomat.com/2024/05/ai-and-elections-lessons-from-south-korea/>; see also The Korea Herald, 'Deepfake risks in election', *The Korea Herald* (online, 23 February 2024) https://www.koreaherald.com/article/3333830; see *Public Official Election Act 1994*.

229 Baker McKenzie, Singapore: Law passed to ban deepfakes during general elections (webpage) https://insightplus.bakermckenzie.com/bm/technology-media-telecommunications_1/singapore-law-passed-to-ban-deepfakes-during-general-elections; Vanessa Lim, Natasha Ganesan & Charmaine Jacob, 'Singapore passes law to ban deepfakes of candidates during elections', *CAN* (online, 15 October 2024) <https://www.channelnewsasia.com/singapore/singapore-ban-deepfakes-general-election-candidates-law-4679781>.

230 See generally Information Resilience & Integrity Symposium, *Countering AI Disinformation: Lessons from Taiwan's 2024 Election Defense Strategies* (Report, 2025).

231 Paul Garvey, 'Inside the social media "dark arts" that fuelled One Nation's shock win', *The Australian* (5 September 2026) <https://www.theaustralian.com.au/inquirer/inside-the-social-media-dark-arts-that-fuelled-one-nations-shock-win/news-story/975fd7e1eca83a449eb22b65901499e7>; Jessica Page, 'Liberal leader cries foul over Labor's campaign linking him to Pauline Hanson', *The West Australian* (6 August 2026) <https://thewest.com.au/politics/state-politics/liberal-leader-cries-foul-over-labors-campaign-linking-him-to-pauline-hanson--c-22685286>.

232 Paul Garvey, 'Inside the social media "dark arts" that fuelled One Nation's shock win', *The Australian* (5 September 2026) <https://www.theaustralian.com.au/inquirer/inside-the-social-media-dark-arts-that-fuelled-one-nations-shock-win/news-story/975fd7e1eca83a449eb22b65901499e7>; Jessica Page, 'Liberal leader cries foul over Labor's campaign linking him to Pauline Hanson', *The West Australian* (6 August 2026) <https://thewest.com.au/politics/state-politics/liberal-leader-cries-foul-over-labors-campaign-linking-him-to-pauline-hanson--c-22685286>.

233 Rosie Lewis and Sarah Ison, 'Social media, bots, misinformation to blame for Labor WA loss: Madeline King', *The Australian* (31 August 2026) <https://www.theaustralian.com.au/nation/politics/social-media-bots-misinformation-to-blame-for-labor-wa-loss-madeleine-king/news-story/831b048c65a380e08d2f71039306d54c>.

234 *Electoral Act 1985* (SA) s 115C.

235 *Electoral Act 1985* (SA) ss 115B(1) & 115C(1)(c).

236 *Electoral Act 1985* (SA) s 115D.

237 Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform' (2021) 44(3) *University of New South Wales Law Journal* 983, 991.

238 Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform' (2021) 44(3) *University of New South Wales Law Journal* 983, 1010-1011.

239 Andrew Ray, 'Disinformation, Deepfakes and Democracies: The Need for Legislative Reform' (2021) 44(3) *University of New South Wales Law Journal* 983, 988.

240 Australian Electoral Commission, 'AEC Completes Recount' (Media Release, 4 June 2025).

241 Jasdev Dhaliwal, 'How to Survive the Deepfake Election with McAfee's 2024 Election AI Toolkit' *McAfee Blog* (Online, 28 October 2024) <https://www.mcafee.com/blogs/internet-security/how-to-survive-the-deepfake-election-with-mcafees-2024-election-ai-toolkit/>.

242 *Criminal Code Act 1995* (Cth) ss 474.174A & 474.17AA.

243 *Criminal Code Act 1995* (Cth) ss 474.174A & 474.17AA.

244 See e.g. Eva Blandis, 'Landmark plea as teen admits creating or altering sexually explicit deepfakes', *ABC News* (online, 15 April 2026) <https://www.abc.net.au/news/2026-04-15/sa-teen-pleads-guilty-over-deepfake-images/106567132>.

245 The case details were not published at the time of writing; See generally Eva Blandis, 'SA Teen Pleads Guilty over Deepfake Images', *ABC News* (Online, 15 April 2026) <https://www.abc.net.au/news/2026-04-15/sa-teen-pleads-guilty-over-deepfake-images/106567132>.

246 *Criminal Code Act 1995* (Cth) ss 474.174A & 474.17AA.

247 *Online Safety Act 2021* (Cth) ss 75, 78 & 83.

248 See *Online Safety Act 2021* (Cth) pt 6, div 3.

249 eSafety, *ImageBased Abuse Scheme: Regulatory Guidance* (February 2024) 5.

250 eSafety, *ImageBased Abuse Scheme: Regulatory Guidance* (February 2024) 11.

251 Noelle Martin, 'Online Safety Regulation of Deepfake Abuse: A Case Study on Australia's eSafety Commissioner' (2025) 34(1) *Griffith Law Review* 23, 33.

252 Or within a longer time specified by eSafety. Note that removal notices can be issued to providers and end users. See *Online Safety Act 2021* (Cth) pt 6 div 3.

253 *Online Safety Act 2021* (Cth) pt 5 & 6.

254 See *eSafety Commissioner v Rotondo (No 4)* [2025] FCA 1191.

255 *Online Safety Act 2021* (Cth) s 7.

256 *Online Safety Act 2021* (Cth) ss 88–91.

257 See e.g., Meta, 'More Speech and Fewer Mistakes' (Media Release, 07 January 2025); See e.g. Jobanny Cabrera, 'Meta Changed Its Speech Rules. Then Threats Against Politicians Skyrocketed', *Wired* (online, 09 June 2026) <https://www.wired.com/story/threats-against-politicians-skyrocketed-after-meta-changed-its-speech-rules/>.

258 Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Documentation and Analysis of Online Abuse Targeting Collective Shout* (Online Hate Prevention Institute, 2026) 5–7, 92–95.

259 Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Documentation and Analysis of Online Abuse Targeting Collective Shout* (Online Hate Prevention Institute, 2026) 101.

260 Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Documentation and Analysis of Online Abuse Targeting Collective Shout* (Online Hate Prevention Institute, 2026) 105.

261 Patrick Scolyer-Gray, *Technology Facilitated Violence Against Women: Documentation and Analysis of Online Abuse Targeting Collective Shout* (Online Hate Prevention Institute, 2026) 33, 70-71.

262 See generally Australian Human Rights Commission, Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts on *the Exposure Draft of the Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* (Submission, 18 August 2023).

263 See generally Australian Human Rights Commission, Submission to the Environment and Communications Legislation Committee on *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024* (Submission, 07 October 2024).

264 See generally Australian Human Rights Commission, Submission to the Environment and Communications Legislation Committee on *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2024* (Submission, 07 October 2024).

265 Digital Industry Group Inc, *Australian Code of Practice on Disinformation and Misinformation* (December 2022).

266 TikTok, *Integrity and Authenticity* (Webpage) <https://www.tiktok.com/safety/en/policies-and-engagement/integrity-authenticity>; TikTok, 'Protecting the integrity of elections', *Transparency Center* (Webpage) <https://www.tiktok.com/transparency/en-us/protecting-elections/>.

267 TikTok, 'Protecting the integrity of elections', *Transparency Center* (Webpage) <https://www.tiktok.com/transparency/en-us/protecting-elections/>.

268 YouTube Help, *Elections misinformation policies* (webpage) <https://support.google.com/youtube/answer/10835034?hl=en>.

269 See generally Irene Khan, *Threats to Freedom of Expression Online in Turbulent Times* (Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN Doc A/80/341, August 2025).

270 Emma Berry, 'Is Misuse of Deepfake Technology Adequately Addressed by Australian Intellectual Property Law, Or Does Australia Need to Introduce a Right of Publicity Tort?' (2025) 170 *Intellectual Property Forum* 7, 8.

271 Emma Berry, 'Is Misuse of Deepfake Technology Adequately Addressed by Australian Intellectual Property Law, Or Does Australia Need to Introduce a Right of Publicity Tort?' (2025) 170 *Intellectual Property Forum* 7, 8.

272 Sabrina Alonso & Darsshana Kannan, *More layers than their pastries: Denmark's deepfake copyright proposal* (Webpage) <https://www.lexology.com/library/detail.aspx?g=d0896e64-8fa1-43c4-857f-fd9adfe40db2>; Sofie Axelsson, *Owning the Self: Denmark's Copyright Turn Against Deepfakes* (Webpage) <https://schjodt.com/news/owning-the-self-denmarks-copyright-turn-against-deepfakes>.

273 Sabrina Alonso & Darsshana Kannan, *More layers than their pastries: Denmark's deepfake copyright proposal* (Webpage) <https://www.lexology.com/library/detail.aspx?g=d0896e64-8fa1-43c4-857f-fd9adfe40db2>; Sofie Axelsson, *Owning the Self: Denmark's Copyright Turn Against Deepfakes* (Webpage) <https://schjodt.com/news/owning-the-self-denmarks-copyright-turn-against-deepfakes>.

274 See *NO FAKES Act of 2025* (US) s 1367.

275 *NO FAKES Act of 2025* (US) s 2(E)(4)(A).

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

277 See Jacqueline So, 'UK Solicitors Regulation Authority investigating criminal law solicitor over TikTok deep fake', *Canadian Lawyer* (online, 21 January 2025) <https://www.canadianlawyermag.com/news/international/uk-solicitors-regulation-authority-investigating-criminal-law-solicitor-over-tiktok-deep-fake/390803>.

278 See Jacqueline So, 'UK Solicitors Regulation Authority investigating criminal law solicitor over TikTok deep fake', *Canadian Lawyer* (online, 21 January 2025) <https://www.canadianlawyermag.com/news/international/uk-solicitors-regulation-authority-investigating-criminal-law-solicitor-over-tiktok-deep-fake/390803>.

279 *Privacy Act 1988* (Cth) sch 2; Attorney-General's Department, *Government Response Privacy Act Review* (Report, 2023) 19.

280 Australian Law Reform Commission, *Serious Invasions of Privacy in the Digital Era* (Summary of Report 123, June 2014) 6 [1.4].

281 See generally *Privacy and Other Legislation Amendment Bill 2024* (Cth).

282 Explanatory Memorandum, Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025.

283 *Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025* (Cth) sch 1.

284 *Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025* (Cth) sch 1, 10.

285 *Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025* (Cth) sch 2, 16-21.

286 See generally *Online Safety and Other Legislation Amendment (My Face, My Rights) Bill 2025* (Cth).

287 Peter Martinez, 'Winnie the Pooh censored in China after President Xi Jinping comparisons', *CBS News* (online, 17 July 2017) <https://www.cbsnews.com/news/winnie-the-pooh-censored-china-president-xi-jinping-comparisons/>; Alan McGuinness, 'China censors Winnie the Pooh social media posts amid Xi criticism', *Sky News* (online, 28 February 2018) https://news.sky.com/story/china-censors-winnie-the-pooh-social-media-posts-amid-xi-criticism-11270677.

288 Peter Martinez, 'Winnie the Pooh censored in China after President Xi Jinping comparisons', *CBS News* (online, 17 July 2017) <https://www.cbsnews.com/news/winnie-the-pooh-censored-china-president-xi-jinping-comparisons/>; Alan McGuinness, 'China censors Winnie the Pooh social media posts amid Xi criticism', *Sky News* (online, 28 February 2018) <https://news.sky.com/story/china-censors-winnie-the-pooh-social-media-posts-amid-xi-criticism-11270677>.

289 See e.g. Department of Industry, Science and Resources, *Safe and Responsible AI in Australian Consultation: Australian Government's Interim Response* (Interim response, 17 January 2024) 20.

290 Carnegie Endowment for International Peace, *Countering Disinformation Effectively: An EvidenceBased Policy Guide* (Report, 2024) 6.

291 *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 (Artificial Intelligence Act)*.

292 *Regulation (EU) 2022/2065* of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending *Directive 2000/31/EC (Digital Services Act)*.

293 which took effect on 2 August 2026.

294 See especially arts 50(2) and 50(4).

295 Provisions on the *Administration of Deep Synthesis Internet Information Services*, the *Measures for the Labelling of Artificial Intelligence Generated Content (Measures)* and the mandatory national standard, *Cybersecurity Content (GB45438—2025) (National Standards)*. The Measures and National Standards came into effect on 1 September 2025.

296 Toby Bond, 'New AI Content Labelling Rules in China: What are they and how do they compare to the EU AI Act?', *Bird & Bird* (20 May 2025) <https://www.twobirds.com/en/insights/2025/new-ai-content-labelling-rules-in-china-what-are-they-and-how-do-they-compare-to-the-eu-ai-act>.

297 Toby Bond, 'New AI Content Labelling Rules in China: What are they and how do they compare to the EU AI Act?', *Bird & Bird* (20 May 2025) <https://www.twobirds.com/en/insights/2025/new-ai-content-labelling-rules-in-china-what-are-they-and-how-do-they-compare-to-the-eu-ai-act>.

298 Proudly Human, *Proudly Human* (Webpage) <https://www.proudlyhuman.org/>.

299 See e.g. Peter Lewis, 'Could AI write this column? In a world of slop-inion, I'm certifying myself human', *The Guardian* (online, 15 April 2026) <https://www.theguardian.com/commentisfree/2026/apr/14/ai-opinion-piece-column-writing-articles-certified-human-writer>;

300 TikTok, 'About AI-generated content', *TikTok Support* (Webpage) <https://www.tiktok.com/support/faq_detail?category=web_posts&id=7636670084747893268>.

301 TikTok, 'About AI-generated content', *TikTok Support* (Webpage) <https://www.tiktok.com/support/faq_detail?category=web_posts&id=7636670084747893268>.

302 See generally Department of Industry, Science and Resources, *Being Clear About AI-Generated Content* (Report, 28 November 2025).

303 Forster et al., *A Multi-Level Strategy for Deepfake Content Moderation under EU Regulation* (Research Paper, 10 July 2025) 7-8.

304 Forster et al., *A Multi-Level Strategy for Deepfake Content Moderation under EU Regulation* (Research Paper, 10 July 2025) 7-8.

305 Forster et al., *A Multi-Level Strategy for Deepfake Content Moderation under EU Regulation* (Research Paper, 10 July 2025) 7-8.

306 See especially Dilrukshi Gamage et al., *Labelling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media* (Conference Paper, CHI Conference on Human Factors in Computing Systems (CHI'25)) 13-15.

307 Dilrukshi Gamage et al., *Labelling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media* (Conference Paper, CHI Conference on Human Factors in Computing Systems (CHI'25)); John Blanchar & Catherine Norris, 'Trump, Twitter, and truth judgments: The effects of 'disputed' tags and political knowledge on the judged truthfulness of election misinformation' (2024) 5(5) *Harvard Kennedy School Misinformation Review* 1.

308 Carnegie Endowment for International Peace, *Countering Disinformation Effectively: An EvidenceBased Policy Guide* (Report, 2024) 6.

309 Carnegie Endowment for International Peace, *Countering Disinformation Effectively: An EvidenceBased Policy Guide* (Report, 2024) 6 & 38.

310 See generally eSafety Commissioner, *Recommender Systems and Algorithms – Full Position Statement* (Position Statement, 11 October 2023).

311 Reddit, *December 2025 Transparency Report* (Report, December 2025).

312 Reddit, *December 2025 Transparency Report* (Report, December 2025).

313 Reddit, *December 2025 Transparency Report* (Report, December 2025).

314 Reddit, *December 2025 Transparency Report* (Report, December 2025).

315 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence* (Interim Report, Commonwealth of Australia, November 2024) 5 [2.5].

316 Select Committee on Adopting Artificial Intelligence (AI), *Adopting Artificial Intelligence* (Interim Report, Commonwealth of Australia, November 2024) 11 [2.21] quoting ANU Tech Policy Design Centre, Submission 68, 4.

317 Nadia Naffi, 'Deepfakes and the Crisis of Knowing', *UNESCO* (Web Page, 1 October 2025) <https://www.unesco.org/en/articles/deepfakes-and-crisis-knowing>.

318 Centre for Emerging Technology and Security, *AIEnabled Influence Operations: Safeguarding Future Elections* (Research Report, 2024) 14; See generally Andrew Guess, Jonathan Nagler & Joshua Tucker, 'Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook' (2019) 5(1) *Science Advances*; See also Center for Countering Digital Hate, 'Belief in conspiracy theories higher among teenagers than adults, as majority of Americans support social media reform, new polling finds' (Online, 16 August 2023) <https://counterhate.com/blog/belief-in-conspiracy-theories-higher-among-teenagers-than-adults-as-majority-of-americans-support-social-media-reform-new-polling-finds/>; Manjul Gupta et al., 'Fake News Believability: The Effects of Political Beliefs and Espoused Cultural Values' (2023) 60(2) *Information & Management*.

319 Centre for Emerging Technology and Security, *AIEnabled Influence Operations: Safeguarding Future Elections* (Research Report, 2024) 17; Carnegie Endowment for International Peace, *Countering Disinformation Effectively* (Report, 2019) 23.

320 See generally Sacha Altay, et al., 'A Survey of Expert Views on Misinformation: Definitions, Determinants, Solutions, and Future of the Field' (2023) 4(4) *Harvard Kennedy School Misinformation Review* 1; Lee Fong Yee et al., 'Social Platforms in the Deepfake Age: Navigating Media Trust through Media Literacy' (2025) *International Journal on Informatics Visualization* 176, 177–178.

321 Carnegie Endowment for International Peace, *Countering Disinformation Effectively* (Report, 2019) 26.

322 Esther Chan, 'First Draft: Prebunking and Avoiding Amplification of Election Disinformation', *BBC* (Online, 2023) <https://www.bbc.co.uk/beyondfakenews/trusted-news-initiative/firstdraftelections2>.

323 Australian Public Service Commission, *Trust and Satisfaction in Australian Democracy* (Report, 2023) 12.

324 Australian Media Literacy Alliance, *Australian Media Literacy Report 2024* (Report, 2024) 15.

325 Australian Media Literacy Alliance, *Australian Media Literacy Report 2024* (Report, 2024) 15.



GPO Box 5218 SYDNEY NSW 2001
Telephone: (02) 9284 9600
National Information Service: 1300 656 419
General enquiries: 1300 369 711
TTY: 1800 620 241
Fax: (02) 9284 9611
Website: **www.humanrights.gov.au**