

Bond University

Bond Law Review

Volume 38

Issue 2

2026

Algorithmic Publicity and the Right to a Fair Trial

Ariss Laghai

Law / Business, Bond University, Law, Kingston University

Follow this and additional works at: <https://blr.scholasticahq.com/>



© Copyright the authors.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivative 4.0 International Licence](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Algorithmic Publicity and the Right to a Fair Trial

ARISS LAGHAI*[◇]

Abstract

This article examines how social media and AI-driven algorithmic publicity affect the right to a fair trial in Australian criminal jury proceedings. It outlines the common law and statutory controls on prejudicial publicity, including sub judice contempt, permanent stays, jury directions and evidentiary discretions, and shows how they operate in leading High Court and appellate decisions and recent law reform work. It then explains how search engines, social media feeds, recommender systems, synthetic media and generative AI, through engagement-optimised ranking and amplification, undermine broadcast-era assumptions about juror exposure. Drawing on empirical research on juror psychology, media effects, and juror internet use, the article evaluates the practical capacity of safeguards such as contempt, stays, adjournment, change of venue, sequestration, judge-alone trials, and directions to preserve impartiality where prejudicial material is persistent, searchable, and personalised. It argues for a robust fair-trial standard paired with a more candid account of digital exposure. Building on developments in the UK, EU, US, China and Estonia, it proposes doctrinal and procedural refinements, narrowly targeted platform duties and court-supervised AI monitoring tools to preserve jury trial and open justice while maintaining credible fair-trial guarantees in an algorithmic information environment.

* Bachelor of Laws / Bachelor of Actuarial Science

[◇] I would like to express my sincere appreciation to Assistant Professor Elizabeth Greene, who generously accepted me to complete this Honours paper under her supervision. Her guidance, insight, and willingness to discuss ideas throughout this project have been invaluable. I am deeply grateful for her support and for the opportunity to work with her. I also extend my thanks to Bond University for providing a supportive and enriching academic environment over the past four years. The opportunities, teaching, and community I have experienced since 2022 have played a significant role in my development as an individual. I am especially grateful to the Abedian Foundation, and in particular to Dr Soheil Abedian AM, whose generosity made my time at Bond University possible. Being awarded the Abedian Foundation Scholarship has been an immense privilege, and I remain deeply thankful for the opportunities it has afforded me at this wonderful university. Finally, I thank my family and friends for their encouragement and support throughout my degree. Their belief in me has made this journey both meaningful and enriching.

Contents

<u>Abstract</u>	27
<u>I. Introduction</u>	3
<u>II. Definitions and Assumptions</u>	5
A. <u>Fair Trial and Impartial Tribunal</u>	5
B. <u>Prejudicial Publicity, <i>Sub Judice</i> Contempt, and Open Justice</u>	6
C. <u>Social Media, Algorithmic Amplification and AI</u>	7
D. <u>Scope of the Inquiry</u>	8
<u>III. Findings and Analysis</u>	9
A. <u>Australian Doctrinal Framework on Prejudicial Publicity</u>	9
B. <u>Algorithmic Publicity: Pathways to Prejudice</u>	13
C. <u>Juror Psychology and Human-Algorithm Dynamics</u>	16
D. <u>Adequacy of Existing Safeguards – Doctrinal Evaluation</u>	20
E. <u>Comparative Regulatory Responses</u>	23
F. <u>Court-Supervised AI as a Targeted Response</u>	26
<u>IV. Conclusion and Recommendations</u>	29
1. <u>Doctrinal and Procedural Refinement</u>	56
2. <u>Targeted Platform Regulation</u>	56
3. <u>Court-Supervised AI and Institutional Capacity</u>	56
<u>Objections and Limits</u>	57
<u>Concluding Reflections</u>	57

I Introduction

This article examines the extent to which social media and AI-driven publicity undermine the right to a fair trial in Australian criminal jury proceedings, and identifies the doctrinal, procedural and regulatory reforms required to safeguard juror impartiality in an algorithmic environment.

The modern criminal trial in Australia assumes that juries can put aside prejudicial publicity and decide a case solely on admissible evidence.¹ This premise underpins doctrines on contempt, stays and jury directions, and presumes that prejudicial material appears in discrete bursts and can be neutralised by judicial instructions.² Conceived in an era of limited, identifiable media, those assumptions sit uneasily in a digital environment where most Australians carry smartphones that deliver algorithmically curated news feeds and AI-powered search results.³

Social media and artificial intelligence have altered pre-trial publicity by curating and ranking content to maximise engagement, so material is repeated, personalised and reinforced through social cues.⁴ Generative models can produce commentary about high-profile cases within minutes. Synthetic media, including deepfakes, make it hard to tell what is real and what is altered.⁵ These tools do not simply increase the volume of publicity, but actively reshape how it is encountered – on platforms where news, commentary and user-generated opinion appear side by side in a feed, consumers can treat opinion as part of the facts of the case, shaping impressions of alleged offending and accused persons long before any jury is empanelled.⁶

Australian courts have responded by applying existing fair-trial safeguards to online publicity, rather than reworking doctrine.⁷ Appellate decisions decline to assume that widespread publicity renders a fair trial impossible and insist that directions, contempt powers and other safeguards will ordinarily suffice.⁸

¹ *Zoneff v The Queen* (2000) 200 CLR 234, 261 [67] (Kirby J).

² Victorian Law Reform Commission, *Contempt of Court: Report* (Report No 40, 2020) [10.11]–[10.12] (‘VLRC’).

³ Deloitte Access Economics, *Mobile Nation 2019: The 5G Future* (Report, 2019) 3.

⁴ Smitha Milli et al, ‘Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media’ (2025) 4(3) PNAS Nexus, 1–2.

⁵ Visual Evidence Lab, *Video’s Day in Court: Advancing Equitable Legal Usage of Visual Technologies and AI* (Report, University of Colorado Boulder, 5 November 2025) 3–4.

⁶ Christine L Ruva and Anthony E Coy, ‘Your Bias Is Rubbing Off on Me: The Impact of Pretrial Publicity and Jury Type on Guilt Decisions, Trial Evidence Interpretation, and Impression Formation’ (2020) 26(1) *Psychology, Public Policy and Law* 22, 23.

⁷ *Glennon v The Queen* [1992] HCA 16; (1992) 173 CLR 592, 603–4, 606–7 (‘Glennon’); *The Queen v The Herald & Weekly Times Pty Ltd & Ors* (Ruling No 2) [2020] VSC 800.

⁸ *DPP v Patterson* [2025] VSC 557; *R v Obeid* [2015] NSWSC 897, [56]–[68].

The analysis begins from the perspective that the fair-trial guarantee is a fundamental standard that must be interpreted in consideration of evolving social and technological conditions.⁹ Repeated exposure to emotive, case-related publicity can increase conviction rates and bias juror evaluations of evidence. Standard judicial instructions to disregard that publicity have only a weak mitigating effect – and sometimes no effect at all.¹⁰ Digital platforms amplify these effects through ranking systems that privilege outrage and social validation, and AI tools condense and personalise narratives about alleged offending.¹¹

The central claim is that existing doctrine remains coherent but increasingly reactive. It presupposes a publicity environment that can be delineated and controlled, and a juror who encounters prejudicial material in discrete episodes. Those assumptions sit uneasily with continuous, personalised feeds and AI-generated narratives.¹²

Against that backdrop, the article's contribution is to test those assumptions against contemporary evidence about juror psychology and digital media, and to bring algorithmic and AI-driven publicity squarely into the frame of fair-trial analysis. It argues that Australian law should not abandon its commitment to jury trials or open justice, but should recalibrate a coherent framework that has become overly dependent on optimistic views of juror resilience. The article advances three main responses. First, it calls for doctrinal and procedural refinement of contempt, stays and jury directions in light of realistic accounts of digital exposure. Second, it proposes narrowly targeted statutory duties on major platforms in relation to clearly prejudicial material linked to pending proceedings. Third, it examines the cautious use of court-supervised AI tools to provide judges with a clearer understanding of the online narratives surrounding specific cases.

Methodologically, the article adopts a doctrinal approach, reading High Court authority, statutory frameworks and law reform reports against contemporary digital conditions, and draws on psychological and media research rather than new empirical data. This core analysis is complemented by selective comparative public-law material from the UK, the EU, the US, China, and Estonia, as well as a structured synthesis of existing empirical work on juror psychology, media effects, and digital information practices.

⁹ *Dietrich v The Queen* (1992) 177 CLR 292, 364 (Gaudron J) ('*Dietrich*').

¹⁰ Lori A Hoetger *et al*, 'The Impact of Pretrial Publicity on Mock Juror and Jury Verdicts: A Meta-Analysis' (2022) 46(2) *Law and Human Behavior* 121.

¹¹ Steve Rathje, Jay J Van Bavel and Sander van der Linden, 'Out-Group Animosity Drives Engagement on Social Media' (2021) 118(26) *Proceedings of the National Academy of Sciences of the United States of America* e2024292118.

¹² Cecilia Anthony Das, Joshua Aston and Haydn Rigby, '#Australianjury#socialmedia—Natural Justice on Trial' (2025) 53(1) *University of Western Australia Law Review* 220, 221–2.

This methodology relies on existing empirical and interdisciplinary literature, rather than new data collection, including psychological studies of juror decision-making, media research on digital ecosystems, and scholarship on AI ethics. Pre-digital empirical studies provide an essential baseline for understanding how jurors encountered publicity in a broadcast-media environment; this article asks how those conclusions must be reassessed considering contemporary, algorithmic publicity.¹³ Comparative material from other jurisdictions is used only to test whether regulatory strategies could sensibly be adapted to Australian conditions.

The article first outlines the concepts and assumptions that frame the inquiry, then analyses the doctrinal position on prejudicial publicity and fair trials before examining how social media and AI influence public narratives and juror attitudes. It then turns to juror psychology, the adequacy of safeguards and comparative regulatory responses, considers the potential role of court-supervised AI monitoring, and concludes with reforms to keep the fair-trial guarantee meaningful in an algorithmic environment.

II Definitions and Assumptions

Clear working definitions are needed before evaluating whether existing safeguards remain fit for purpose.

A *Fair Trial and Impartial Tribunal*

In this article, the term ‘fair trial’ refers to the procedural guarantees that must be present for a criminal conviction to be regarded as legitimate.¹⁴ At its core, fairness requires an independent and impartial court, and a verdict based solely on admissible evidence that has been presented and tested according to law, rather than influenced by external pressure or undisclosed material.¹⁵ A fair trial in this sense is not confined to what happens inside the courtroom.¹⁶ It extends to the conditions in which the hearing takes place, including the information environment to which jurors are exposed.¹⁷ When the information environment makes it unrealistic to expect jurors to set aside prejudicial material, or exerts pressure on courts to reflect public anger rather than

¹³ Nancy M Steblay *et al*, ‘The Effects of Pretrial Publicity on Juror Verdicts: A Meta-Analytic Review’ (1999) 23(2) *Law and Human Behavior* 219.

¹⁴ *Murphy v The Queen* (1989) 167 CLR 94, 98–9 (Mason CJ and Toohey J) (‘*Murphy*’).

¹⁵ *International Covenant on Civil and Political Rights*, opened for signature 16 December 1966, 999 UNTS 171 (entered into force 23 March 1976) art 14(1).

¹⁶ VLRC, *Contempt of Court: Report* (n 2) 1–2 [2].

¹⁷ Jemma Holt and Terese Henning, *Jurors, Social Media and the Right of an Accused to a Fair Trial* (Issues Paper No 30, Tasmania Law Reform Institute, 2020) 10.

legal principle, fairness is compromised, even if the proceeding is carefully managed on its face.¹⁸

‘Impartial tribunal’ is used in a practical, trial-level sense and denotes a court constituted according to law, before which the decision-maker has no unacceptable predisposition for or against the accused.¹⁹ For juries, this includes an expectation that jurors will approach the case with an open mind, follow the judge's directions, and set aside any personal knowledge or prior exposure to publicity.²⁰ For trial judges, it includes both actual impartiality and the appearance of impartiality.²¹ Australian authority repeatedly links public confidence in the administration of justice to the perception that courts decide cases based on evidence and law, rather than popular sentiment or media pressure.²²

B *Prejudicial Publicity, Sub Judice Contempt, and Open Justice*

‘Prejudicial publicity’ comprises communications about pending or potential criminal proceedings that risk influencing attitudes towards the accused or the outcome. Such publicity invites the public to treat an accused as guilty before a verdict has been reached. Empirical research supports these concerns: a meta-analysis found that negative pre-trial publicity increases guilty verdicts among individual jurors and deliberating juries.²³

‘*Sub judice* contempt’ is the law’s primary tool for controlling prejudicial publicity. This principle was recently affirmed in the high-profile matter of *Lehrmann v Network Ten Pty Limited*, where the Federal Court expressly acknowledged *sub judice* contempt as a safeguard for the administration of justice.²⁴ Across Australia, the publication must have a real and definite tendency to prejudice proceedings.²⁵ This doctrine targets identifiable acts of publication and, by restricting such material, protects the administration of justice, functioning as an exception to the principle of free expression and open justice.²⁶

¹⁸ *Wainohu v New South Wales* (2011) 243 CLR 181.

¹⁹ *Ebner v Official Trustee in Bankruptcy* [2000] HCA 63; (2000) 205 CLR 337, 345 [6].

²⁰ *Murphy* (n 14).

²¹ *Johnson v Johnson* [2000] HCA 48; (2000) 201 CLR 488, 492 [11].

²² *Webb v The Queen* (1994) 181 CLR 41, 52–3 (Mason CJ and McHugh J).

²³ *Lori A Hoetger et al* (n 10) 132–3.

²⁴ *Lehrmann v Network Ten Pty Limited (Trial Judgment)* [2024] FCA 369, [1038] (*Lehrmann*).

²⁵ *Hinch v Attorney-General* (Vic) (1987) 164 CLR 15, 34 (Wilson J), 70 (Toohey J), 88 (Gaudron J) (*‘Hinch’*).

²⁶ VLRC, *Contempt of Court: Report* (n 2) [10.1]–[10.3].

‘Open justice’ provides that court proceedings should be conducted in public and subject to scrutiny.²⁷ The Australian Law Reform Commission calls it a fundamental attribute of a fair trial: the administration of justice must take place in open court, and proceedings should be subject to public and professional scrutiny.²⁸ It manifests through open hearings, public communication of evidence and encouragement of fair and accurate reporting.²⁹ Yet the principle is not absolute; courts may limit publicity where necessary for the proper administration of justice, national security or the protection of vulnerable witnesses.³⁰

C *Social Media, Algorithmic Amplification and AI*

‘Social media’ is used here to encompass platforms that host user-generated content and facilitate interaction between users, including Facebook, Instagram, X, TikTok, and YouTube. They provide low-friction channels for ordinary users, advocacy groups, influencers and news outlets to circulate narratives about alleged offending.³¹ Content is typically delivered through ranked feeds rather than simple chronological lists, so that some stories are repeatedly surfaced, and others remain effectively invisible.³²

‘Algorithmic amplification’ describes how recommender systems select and prioritise content based on engagement metrics and inferred preferences.³³ Randomised experiments indicate that engagement-based ranking amplifies emotionally charged, out-group-hostile content and does not necessarily improve users’ satisfaction with the material they are shown.³⁴ Such systems privilege outrage and social validation, so posts framing an accused as dangerous or morally reprehensible are more likely to be surfaced repeatedly than measured reporting of the

²⁷ *Commissioner of the Australian Federal Police v Zhao* (2015) 316 ALR 378, [44] (French CJ, Hayne, Kiefel, Bell and Keane JJ).

²⁸ Australian Law Reform Commission, *Traditional Rights and Freedoms—Encroachments by Commonwealth Laws* (Report No 129, 2015) 247 [8.53].

²⁹ Jason Bosland and Ashleigh Bagnall, ‘An Empirical Analysis of Suppression Orders in the Victorian Courts: 2008-12’ (2013) 35 *Sydney Law Review* 674.

³⁰ *John Fairfax Group v Local Court of NSW* (1991) 36 NSWLR 131, 141.

³¹ VLRC, *Contempt of Court: Report* (n 2) [13.5].

³² Motahhare Eslami *et al*, “‘I Always Assumed That I Wasn’t Really That Close to Her’: Reasoning about Invisible Algorithms in the News Feed” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Association for Computing Machinery, 2015).

³³ eSafety Commissioner, *Recommender systems and algorithms* (Position Statement, December 2022) 2–3, 6, 9.

³⁴ Smitha Milli *et al* (n 4).

legal process.³⁵ Over time, these dynamics facilitate personalised echo chambers and continuous exposure to prejudicial narratives.³⁶

‘Artificial intelligence’ is used narrowly to refer to generative models, synthetic media and AI-assisted search and summarisation tools that shape publicity around criminal proceedings. This article also uses the term ‘algorithmic publicity’ to refer to those forms of publicity surrounding criminal proceedings that are structured by engagement-optimised feeds, recommender systems and generative AI tools across these platforms.

D *Scope of the Inquiry*

The argument that follows is confined to serious criminal jury trials in Australia. It does not deal with civil proceedings, tribunals or summary matters, despite similar dynamics being present in those contexts. Jury trials are chosen as the focus because they are the setting in which prejudicial publicity has traditionally been perceived as most acute and in which the fairness of the process depends heavily on the attitudes and behaviour of lay decision-makers drawn from the community.³⁷

The temporal focus is on publicity occurring before and during the trial.³⁸ Post-verdict commentary principally raises questions about public confidence in the criminal justice system, rather than the fairness of a trial that has already concluded. An exception arises during the appeal period, however, where publicity following a verdict may still engage fair-trial concerns – particularly where an appeal is pending and a retrial may be ordered.³⁹ Where a conviction is overturned on appeal, and a retrial is ordered, however, earlier reporting and commentary – including discussion of the first verdict – form part of the pre-trial publicity for the fresh proceedings and may again engage fair-trial concerns.⁴⁰ Within the pre-trial and trial window, the article is interested in both high-profile spikes of attention around events and the slower, background accumulation of narratives about crime and punishment to which potential jurors are exposed in their everyday digital lives.

³⁵ Steve Rathje *et al* (n 11).

³⁶ Eytan Bakshy, Solomon Messing and Lada A Adamic, ‘Exposure to Ideologically Diverse News and Opinion on Facebook’ (2015) 348 *Science* 1130, 1130–2.

³⁷ Michael Chesterman, Janet Chan and Shelley Hampton, *Managing Prejudicial Publicity: An Empirical Study of Criminal Jury Trials in New South Wales* (Law and Justice Foundation of New South Wales, 2001) 10–11.

³⁸ Dietrich (n 9) 300 (Mason CJ and McHugh J).

³⁹ *Criminal Code 1899* (Qld) s 671 (*‘Criminal Code’*).

⁴⁰ *Glennon* (n 7).

III Findings and Analysis

A *Australian Doctrinal Framework on Prejudicial Publicity*

This section argues that Australian courts have developed a doctrinal framework that treats prejudicial publicity as a serious but ultimately manageable threat to the fairness of a criminal trial. The underlying assumption is that existing tools will usually allow proceedings to continue without interruption.⁴¹ Stays, venue changes, and sequestration operate as safety valves when contamination is deemed to be acute.⁴² Jury directions are treated as the primary means by which any residual risk is cured.⁴³ Read together, the leading authorities display steady confidence in lay decision makers and a reluctance to accept that media coverage, even intense coverage, will ordinarily render a trial unfair.⁴⁴

Hinch v Attorney-General (Vic) is the starting point for modern *sub judice* doctrine.⁴⁵ The Court upheld findings of contempt against radio broadcasts that named an accused and set out his prior convictions while fresh charges were pending, holding that they had a ‘real and definite tendency’ to prejudice the administration of justice.⁴⁶ This standard indicates that contempt is concerned not only with deliberate interference, but also with publications that, viewed objectively, are apt to influence jurors or otherwise distort the process, while excluding merely speculative or remote risks.⁴⁷ *Hinch* therefore illustrates both the protective and limiting dimensions of *sub judice* law: it recognises that publicity portraying an accused as already guilty strikes at the presumption of innocence, yet preserves space for robust reporting by requiring a meaningful likelihood of prejudice.⁴⁸ That balance frames *sub judice* contempt as a tool for mediating between fair-trial rights, open justice and free-speech values, rather than a vehicle for broad press control.⁴⁹

The modern principles governing permanent stays were established in *Jago v District Court (NSW)*, where the High Court located the stay power in the court’s inherent jurisdiction to prevent an abuse of

⁴¹ *Dupas v The Queen* [2010] HCA 20; (2010) 241 CLR 237, 248–9 [29], 250–1 [36] (‘*Dupas*’); *Glennon* (n 7).

⁴² *Dupas* (n 41).

⁴³ *Ibid.*

⁴⁴ *Glennon* (n 7); *Dupas* (n 41).

⁴⁵ *Hinch* (n 25).

⁴⁶ *Ibid.*

⁴⁷ *Lehrmann* (n 24); New South Wales Law Reform Commission, Contempt by Publication (Report No 100, 2003) 48–9 [4.7]–[4.10] (‘NSWLRC’).

⁴⁸ *Hinch* (n 25); *Lehrmann* (n 24) [1038]–[1039].

⁴⁹ NSWLRC, Contempt by Publication (n 47) 185–8 [8.7]–[8.11].

process.⁵⁰ It insisted that a permanent stay is exceptional, justified only where the continuation of proceedings would necessarily result in an unfair trial.⁵¹ A fair trial does not mean a perfect trial: some imperfection is tolerated so long as the accused retains a real opportunity to meet the Crown's case.⁵² That understanding underpins later publicity cases, which treat external influences as matters to be managed through process unless they make fairness unattainable in fact.⁵³

Where publicity has already occurred, the focus shifts from restraining publishers to managing their impact on proceedings.⁵⁴ *Glennon v The Queen* involved a retrial conducted amid extensive publicity about an earlier conviction and other allegations, including material that had previously resulted in a finding of *sub judice* contempt in *Hinch*.⁵⁵ The accused argued that no jury could now be impartial and sought a permanent stay.⁵⁶ The High Court rejected that contention and reiterated that a stay is an exceptional remedy, to be granted only where no directions or procedural measures could realistically secure fairness.⁵⁷ Rather than halting proceedings altogether, the Court expressed a preference for proceeding to trial with careful directions, juror questioning and targeted discharge of individual jurors where necessary.⁵⁸ The Court drew a line between *ex ante* control of publicity through contempt and *ex post* management of its effects through the trial process, warning against using the stay jurisdiction as an indirect way to regulate the media.⁵⁹ The question is not whether the press behaved responsibly, but whether, in light of what has been published, a fair trial is still possible.⁶⁰

The High Court revisited these themes in *Dupas v The Queen*, a prosecution of a notorious serial murderer whose crimes had attracted sustained and sensational publicity.⁶¹ *Dupas* argued that notoriety made it impossible to empanel an impartial jury and that directions could not

⁵⁰ *Jago v District Court of New South Wales* [1989] HCA 46; (1989) 168 CLR 23, 28–33 (Mason CJ) ('*Jago*').

⁵¹ *Ibid.*

⁵² *Ibid.*

⁵³ *Glennon* (n 7).

⁵⁴ *Dupas* (n 41).

⁵⁵ *Glennon* (n 7); *Hinch* (n 25).

⁵⁶ *Glennon* (n 7).

⁵⁷ *Ibid.*

⁵⁸ *Dupas* (n 41).

⁵⁹ *Ibid.*

⁶⁰ *Glennon* (n 7).

⁶¹ *Dupas* (n 41).

neutralise the effects of coverage.⁶² The Court again declined to treat notoriety as a bar to trial. Fairness did not require jurors to be wholly ignorant of the case; the critical question was whether, after directions, they could decide the matter on the evidence.⁶³ *Dupas* is often regarded as the pinnacle of judicial confidence in jury resilience.⁶⁴ The Court acknowledged the extremity of the publicity yet placed the onus on the accused to demonstrate that no combination of safeguards could ensure fairness, and reasoned *a priori* largely, without engaging with empirical research on media effects or juror psychology.⁶⁵

Queensland-specific scholarship and institutional responses have raised concerns about a more solid foundation. Isaac Frawley Buckley argued in 2013 that pre-trial publicity and social media pose challenges to protecting impartiality, highlighting how online commentary, sharing, and archiving can entrench prejudicial narratives long after traditional media cycles have moved on.⁶⁶ He suggested that existing contempt, stay and jury-direction doctrines, developed for a broadcast landscape, risked underestimating both the persistence of digital material and the difficulty of preventing jurors from encountering it in their ordinary online lives.⁶⁷ Around the same time, Blackham and Williams' study of Australian courts and social media recorded judicial officers and court staff expressing concern that the collaborative nature of platforms sits uneasily with traditional adjudicative processes, and that the decentralised, user-driven character of online publication makes breaches of suppression orders and contemptuous commentary harder to detect and sanction.⁶⁸

Alongside these authorities, evidentiary discretions provide an internal safeguard within the trial itself. Sections 135 to 137 of the *Evidence Act 1995* (Cth) ('*Evidence Act*') empower courts to exclude evidence where the danger of unfair prejudice outweighs its probative value, or where there is a risk of misleading or confusing the jury.⁶⁹ Equivalent discretions exist in Queensland and other non-uniform Evidence Act jurisdictions, including under section 130 of the *Evidence Act 1977* (Qld).⁷⁰ These provisions are routinely used to prevent jurors

62 Ibid.

63 Ibid.

64 Ibid.

65 Ibid.

66 Isaac Frawley Buckley, 'Pre-trial publicity, social media and the "fair trial": protecting impartiality in the Queensland criminal justice system' (2013) 33(1) *Queensland Lawyer* 38.

67 Ibid.

68 Alysia Blackham and George Williams, 'Social Media and the Courts' (2013) 38(2) *Alternative Law Journal* 82.

69 *Evidence Act 1995* (Cth) ss 135–137 ('*Evidence Act*').

70 *Evidence Act 1977* (Qld) s 130.

from hearing about prior misconduct, marginally relevant allegations, or unduly inflammatory material, and they give concrete effect to the idea that jurors should decide cases based on logically probative material rather than on character attacks or emotive details. However, the discretions are confined to the evidentiary record. They operate on what the parties seek to tender in court, not on the impressions jurors may already have formed through news reports, commentary threads or social media feeds. A judge may exclude a graphic image or an account of past offending from the prosecution's case, but cannot retrospectively excise a deepfake video circulated months earlier or a sequence of online posts that depicted the accused as already guilty.⁷¹ In that sense, the statutory discretions and the common law doctrines sit alongside each other but address different aspects of the problem.

Law reform bodies have recognised that this framework is under strain.⁷² The Australian Law Reform Commission's Contempt Report emphasised that contempt law must balance freedom of expression, open justice and the right to a fair trial.⁷³ It recommended a statutory restatement to improve clarity and predictability. The Victorian Law Reform Commission's Contempt of Court report likewise identified online publishing and social media as having changed the way people access and share information.⁷⁴ It concluded that the law of contempt has not kept up and recommended more precise definitions of *sub judice* contempt and modernised procedures for enforcing publication restrictions, including better systems for notifying and monitoring suppression orders. New South Wales and other jurisdictions have undertaken related reviews of contempt by publication and jury directions, noting the difficulty of shielding jurors from internet-based material and the limits of relying on self-reporting about exposure.⁷⁵ These reports accept the basic structure of existing doctrines and the importance of fair trial protections, but acknowledge that enforcement is becoming harder as publicity migrates to digital platforms.

Taken together, these materials outline a framework that prioritises protecting fair trials while fostering robust public discussion. The jurisdiction to stay proceedings is reserved for exceptional cases in which fairness cannot be maintained credibly.⁷⁶ Jury directions, venue changes, and sequestration are typically considered sufficient responses to most instances of pre-trial and mid-trial publicity, with evidentiary

⁷¹ Visual Evidence Lab (n 5) 3–4.

⁷² VLRC, *Contempt of Court: Report* (n 2) [10.1]–[10.3], [13.5].

⁷³ Australian Law Reform Commission, *Contempt* (Report No 35, 1987).

⁷⁴ VLRC, *Contempt of Court: Report* (n 2).

⁷⁵ NSWLRC, *Contempt by Publication* (n 47).

⁷⁶ *Jago* (n 50).

discretion reinforcing the focus on probative material.⁷⁷ Law reform commissions have urged statutory restatement and modernisation, particularly in light of online publishing, but the underlying assumptions about juror behaviour and the manageability of publicity remain largely intact.⁷⁸

B *Algorithmic Publicity: Pathways to Prejudice*

The shift from broadcast media to digital platforms has not only increased the volume of publicity surrounding criminal cases but it has also altered how that publicity is produced, selected, and disseminated.⁷⁹ The risk is no longer confined to a chance newspaper article or vaguely remembered television report.⁸⁰ Potential jurors now inhabit information environments in which content linked to alleged offending is continuously ranked, repeated and reframed by systems designed to maximise engagement.⁸¹ Those design choices create new pathways for prejudice.

Most major platforms present content through curated feeds rather than chronological lists.⁸² Recommender systems draw on past clicks, watch time, follows and social connections to decide what appears next on a user's screen, and material that provokes strong reactions tends to be amplified.⁸³ Posts that frame an accused as dangerous, morally reprehensible or emblematic of a wider social problem are often more engaging than careful summaries of legal process, and so are more likely to be surfaced repeatedly.⁸⁴ Over time, a user may encounter numerous fragments – headlines, clips, comment threads – each reinforcing the same picture and normalising a particular view of the accused.⁸⁵

Once a post about an alleged offence begins to attract interaction, platform algorithms can quickly extend its reach beyond the original audience.⁸⁶ Influencers, advocacy accounts and mainstream outlets

⁷⁷ New South Wales Public Defenders, *Dealing with Prejudicial and Adverse Publicity* (Paper, New South Wales Public Defenders, undated) 3–5.

⁷⁸ VLRC, *Contempt of Court: Report* (n 2); NSWLRC, *Contempt by Publication* (n 47); Australian Law Reform Commission, *Contempt* (n 73).

⁷⁹ Deloitte Access Economics (n 3) 3.

⁸⁰ News & Media Research Centre, *Digital News Report: Australia 2023* (Report, University of Canberra, 2023) 8–10.

⁸¹ Motahhare Eslami *et al* (n 32).

⁸² Eytan Bakshy *et al* (n 36) 1130.

⁸³ Smitha Milli *et al* (n 4).

⁸⁴ William J Brady *et al*, 'Emotion Shapes the Diffusion of Moralized Content in Social Networks' (2017) 114(28) *Proceedings of the National Academy of Sciences* 7313.

⁸⁵ Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin, 2011).

⁸⁶ Cass R Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press, 2017).

may adapt the story, producing overlapping waves of coverage.⁸⁷ Echo-chamber effects occur when individuals who already hold strong attitudes towards crime are presented with similar reactions from others, creating the impression of a broad consensus.⁸⁸ For a future juror situated within that network, the case may already feel familiar and morally resolved before any indictment is read.

Techniques for fabricating images, audio and video are now widely accessible.⁸⁹ In the context of criminal narratives, deepfakes can be used to create apparently authentic recordings of an accused confessing, re-enacting the alleged offence or otherwise reinforcing a particular characterisation.⁹⁰ There is emerging evidence that even attentive media users, including prospective jurors, may struggle to distinguish authentic from fabricated footage encountered online.⁹¹ In that case, there is a risk that prejudicial material will enter their mental picture of the case and that unwarranted scepticism will develop towards legitimate exhibits tendered at trial.⁹² Although some deepfake uses are now criminalised, a fabricated confession that circulates for months before a homicide trial may never be the subject of specific charges, yet it can shape community attitudes and juror expectations in ways that ordinary contempt and suppression tools struggle to reach.⁹³

Generative AI systems that produce text and images also function as new commentators.⁹⁴ When prompted with questions about a high-profile case, chatbots and AI-enhanced search tools may synthesise whatever information is available, accurate or otherwise, and present it as a coherent narrative.⁹⁵ A potential juror who asks a chatbot about a widely discussed incident is not working through a set of identifiable sources; they are receiving a single, smoothed story, and the

⁸⁷ Andrew Chadwick, *The Hybrid Media System: Politics and Power* (2nd ed, Oxford University Press, 2017).

⁸⁸ Zeynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (Yale University Press, 2017).

⁸⁹ *Visual Evidence Lab* (n 5) 3–4.

⁹⁰ Robert Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 *California Law Review* 1753.

⁹¹ Alexander Diel *et al*, 'Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers' (2024) 16 *Computers in Human Behavior Reports* 100538.

⁹² Chesney and Citron (n 90).

⁹³ *Visual Evidence Lab* (n 5).

⁹⁴ Felix M Simon, *Artificial Intelligence in the News: How AI Retools, Rationalizes, and Reshapes Journalism and the Public Arena* (Tow Center for Digital Journalism, 2024).

⁹⁵ Andreas Jungherr and Ralph Schroeder, 'Artificial Intelligence and the Public Arena' (2023) 33(2–3) *Communication Theory* 164.

conversational tone can make the impressions formed remarkably durable.⁹⁶

These systems ensure that users who show interest in crime-related content are steadily offered more of the same.⁹⁷ Recommendation tools and automatic summaries help organise this material into themes – such as youth violence, lenient judges, and repeat offenders – so that individual cases become examples within those themes rather than isolated events.⁹⁸ When one of those cases later comes before a jury, some members of the panel may approach it not as a blank slate but as another chapter in a story they have been seeing for months or years.⁹⁹

Juror search behaviour sits at the intersection of these dynamics and traditional concerns about extra-curial information.¹⁰⁰ Courts routinely instruct jurors not to conduct their own research, not to search the internet for information about the accused, and not to discuss the case on social media.¹⁰¹ Yet, breaches of these instructions have been documented since the widespread adoption of internet access.¹⁰² The contemporary environment changes the ease with which even a fleeting curiosity can trigger a chain of exposure. Typing a name or case label into a search bar does not simply retrieve a single neutral document; it produces ranked results, suggested questions, related videos, and links to commentary.¹⁰³ Algorithms then treat that search as a signal of interest. Even a juror who obeys directions not to ‘look anything up’ may still have case-related posts pushed into their ordinary social-media feeds whenever they open the apps they use every day.¹⁰⁴

-
- ⁹⁶ Timothy Hannigan, Ian P McCarthy and Andre Spicer, ‘Beware of Botshit: How to Manage the Epistemic Risks of Generative Chatbots’ (Working Paper, 28 December 2023) 1–2 <<https://ssrn.com/abstract=4678265>>.
- ⁹⁷ Chris Meserole, ‘How Do Recommender Systems Work on Digital Platforms? Social Media Recommendation Algorithms’ (Article, Brookings Institution, 21 September 2022) <<https://www.brookings.edu/articles/how-do-recommender-systems-work-on-digital-platforms-social-media-recommendation-algorithms/>>.
- ⁹⁸ Kate Goddard, Abdul Roudsari and Jeremy C Wyatt, ‘Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators’ (2012) 19(1) *Journal of the American Medical Informatics Association* 121, 121–2.
- ⁹⁹ eSafety Commissioner, ‘An Unfair Fight – How Algorithms Are Shaping Our Adolescents’ (Online Safety Advisory, 17 April 2025) <<https://www.esafety.gov.au/newsroom/blogs/an-unfair-fight-how-algorithms-are-shaping-our-adolescents>>.
- ¹⁰⁰ Ruva and Coy (n 6) 23–4.
- ¹⁰¹ Lori A Hoetger *et al* (n 10) 132–3.
- ¹⁰² Robert Booth, ‘Violence on Social Media Making Teenagers Afraid to Go Out, Study Finds’ (*The Guardian*, Online, 25 November 2024) <<https://www.theguardian.com/media/2024/nov/25/violence-on-social-media-making-teenagers-afraid-to-go-out-study-finds>>.
- ¹⁰³ Australian Competition and Consumer Commission, *Digital Platform Services Inquiry: Interim Report 9: Revisiting General Search Services* (Report, September 2024) 10–13.
- ¹⁰⁴ Senate Economics References Committee, *Digital Platforms: Improving the Transparency of Algorithms and Artificial Intelligence* (Commonwealth of Australia, 2023) 124–6.

From the perspective of evidence law, these developments create both familiar and novel challenges. Questions of authentication, reliability, and chain of custody arise in an intensified form when images and recordings can be fabricated with minimal technical skill.¹⁰⁵ Courts can and do respond by demanding greater forensic support for audiovisual exhibits and by reminding juries that online material is not evidence unless it is proven and admitted.¹⁰⁶ Yet much of the prejudicial force of algorithmic publicity operates upstream of the evidential process. Deepfakes and AI-generated narratives can influence what potential jurors believe long before any evidence is presented, and without leaving a clear trail that can be confronted in court.¹⁰⁷

The pathways described here – curated feeds, viral amplification, synthetic media, AI-generated narratives and routine search behaviour – do not operate in isolation.¹⁰⁸ They interact to form a dense informational backdrop against which criminal proceedings unfold. The leading authorities on prejudicial publicity were crafted on the assumption that exposure is episodic, that its sources are identifiable and that its influence can be countered by instruction.¹⁰⁹ The digital environment challenges each of those assumptions. Prejudice can be formed gradually through repeated low-level encounters; the origin of particular impressions may be impossible to reconstruct; and the content itself may blend facts, opinion and fabrication in ways that are difficult to disentangle.¹¹⁰ Those features do not establish that fair trials are now unattainable, but they do suggest that any assessment of risk that ignores algorithmic publicity is incomplete.

C *Juror Psychology and Human-Algorithm Dynamics*

The legal framework assumes a juror who may encounter publicity but can set it aside when instructed, and who will generally comply with

¹⁰⁵ Rebecca A Delfino, 'Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery' (2023) 74 *Hastings Law Journal* 293, 298–301.

¹⁰⁶ David Cole, 'Technology and Crime: How Digital Evidence Is Changing Criminal Trials' (Web Page, David Cole Lawyer, 3 July 2022) <<https://www.davidcoleglawyer.com.au/technology-and-crime-digital-evidence-changing-criminal-trials.html>>.

¹⁰⁷ Hannah Smith and Katherine Mansted, *Weaponised Deep Fakes: National Security and Democracy* (Australian Strategic Policy Institute, 29 April 2020).

¹⁰⁸ N Boroda, 'What Are SERP Features? 18 Google Search Features to Know' (Web Page, Semrush, 8 January 2022) <<https://www.semrush.com/blog/serp-features/>>.

¹⁰⁹ *Glennon* (n 7); *Hinch* (n 25); *Dupas* (n 41).

¹¹⁰ Tasmania Law Reform Institute, *Jurors, Social Media and the Right of an Accused to a Fair Trial* (Final Report No 30, January 2020) 72–4, 79–83 ('TLRF').

directions not to seek additional information.¹¹¹ Empirical work on juries and pre-trial publicity presents a more complicated account.¹¹²

Mock-jury experiments and survey studies consistently find that pre-trial publicity can shift both verdict preferences and assessments of guilt.¹¹³ Early research found that exposure to negative publicity increased conviction rates compared to control groups and that standard judicial instructions had little effect in mitigating this effect.¹¹⁴ Studies of actual and simulated juries indicate that pre-existing attitudes toward crime and punishment, as well as prior knowledge of a case, are strong predictors of how evidence is interpreted.¹¹⁵ Chesterman, Chan and Hampton's empirical study of New South Wales jury trials provides a pre-digital Australian baseline.¹¹⁶ Many jurors reported prior media exposure, a smaller group continued to recall that material during deliberations, and most judges regarded directions as effective while still recognising some residual concerns.¹¹⁷

Assessments of remedial measures reach similarly cautious conclusions: reviews of the empirical literature suggest that instructions warning jurors to disregard publicity or inadmissible material have, at best, modest and uneven effects.¹¹⁸ Adjournments and changes of venue can reduce the salience of particular stories, but they do not remove longer-term impressions of widely discussed cases.¹¹⁹ Australian research on prejudicial publicity indicates that the recall of media coverage fades over time but does not disappear, and that some jurors continue to rely on extrajudicial material despite clear directions.¹²⁰ Burd and Horan's survey of Australian judges and practitioners found that concerns about jurors' internet use were no longer hypothetical: respondents reported both deliberate online research and 'incidental' exposure, as jurors 'stumbled across' case-related material when

-
- ¹¹¹ Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy-making* (Council of Europe, Report DGI(2017)09, 2017).
- ¹¹² Rebecca McEwen and John Eldridge, 'Judges, Juries and Prejudicial Publicity: Lessons from Empirical Legal Scholarship' (2016) 41(2) *Alternative Law Journal* 110.
- ¹¹³ Lori A Hoetger *et al* (n 10).
- ¹¹⁴ Geoffrey P Kramer, Norbert L Kerr and John S Carroll, 'Pretrial Publicity, Judicial Remedies, and Jury Bias' (1990) 14(5) *Law and Human Behavior* 409.
- ¹¹⁵ Tarika Daftary-Kapur, Richard Dumas and Steven D Penrod, 'Jury Decision-Making Biases and Methods to Counter Them' (2010) 15(1) *Legal and Criminological Psychology* 133, 137–40.
- ¹¹⁶ Michael Chesterman *et al* (n 37).
- ¹¹⁷ *Ibid.*
- ¹¹⁸ Australian Law Reform Commission, *Uniform Evidence Law* (Report No 102, 2005) [18.8]–[18.10].
- ¹¹⁹ *Glennon* (n 7); *Dupas* (n 41).
- ¹²⁰ Michael Chesterman *et al* (n 37).

platform feeds surfaced trending stories unprompted.¹²¹ Some experimental work suggests that strong instructions to disregard or avoid specific evidence can backfire, with jurors paying more attention to the very material they are told to ignore – an effect explained by psychological reactance and Wegner’s ‘ironic processes’ of mental control.¹²² These findings do not demonstrate that fair trials cannot occur; yet they cast doubt on the assumption that procedural measures can realistically remove the influence of earlier publicity.

Digital technologies complicate this further by altering both the pattern of exposure and the cognitive context in which information is processed. Australian and international work on jurors and social media has documented instances in which jurors have conducted online research, checked news sites or used social platforms during trials, despite explicit instructions not to do so.¹²³ Legal responses – stronger warnings, more explicit written directions and occasional discharge or contempt proceedings – address obvious misconduct but do little to counter what jurors have already seen in their ordinary online lives.

Engagement-optimised algorithms shape the broader media environment in which that behaviour occurs.¹²⁴ Studies of social networks indicate that posts expressing moral outrage or hostility towards out-groups attract more engagement and are preferentially amplified. Research on misinformation diffusion suggests that emotionally charged content spreads more widely and quickly than neutral corrections.¹²⁵ Social feedback – likes, shares and comments – reinforces this pattern by rewarding strong views, so punitive or accusatory narratives about crime can come to feel both common and socially endorsed.¹²⁶

Limited digital verification skills sit alongside these dynamics. Australian studies suggest that many adults have only a basic ability to assess the reliability of online material, despite regularly encountering

-
- ¹²¹ Roxanne Burd and Jacqueline Horan, ‘Protecting the Right to a Fair Trial in the 21st Century: Has Trial by Jury Been Caught in the World Wide Web?’ (2012) 36(2) *Criminal Law Journal* 103, 111–13.
- ¹²² Daniel M Wegner, ‘Ironic Processes of Mental Control’ (1994) 101(1) *Psychological Review* 34, 34–6.
- ¹²³ Thaddeus A Hoffmeister, ‘Google, Gadgets, and Guilt: Juror Misconduct in the Digital Age’ (2012) 83 *University of Colorado Law Review* 409, 414–20.
- ¹²⁴ Na Cai, Shuhong Gao and Jinzhe Yan, ‘How the Communication Style of Chatbots Influences Consumers’ Satisfaction, Trust, and Engagement in the Context of Service Failure’ (2024) 11 *Humanities and Social Sciences Communications* 687.
- ¹²⁵ Donghee Shin and Emily Y Shin, ‘Cascading Falsehoods: Mapping the Diffusion of Misinformation in Algorithmic Environments’ (2025) *AI & Society* (advance).
- ¹²⁶ Yuchen Wu, Mingduo Zhao and John F Canny, ‘Beyond Likes: How Normative Feedback Complements Engagement Signals on Social Media’ (Preprint, 14 May 2025) <<https://arxiv.org/abs/2505.09583>>.

false or misleading content.¹²⁷ At the same time, younger Australians are described as the most connected generation, with nearly universal access to personal devices and a high level of integration of social media into their daily lives.¹²⁸ Many will be practised at navigating platforms, but less adept at distinguishing reliable reporting from speculation, commentary or manipulated press. Jurors have always brought personal experience to the task of decision-making; the concern here is not that they have impressions, but that those impressions are increasingly formed through curated feeds that reward sensational or polarised crime coverage.¹²⁹ When they later sit on a jury, impressions formed through years of such exposure inevitably inform their sense of who tends to offend, how the system usually responds, and what outcomes they consider fair.¹³⁰

Viewed in this way, engagement-based algorithms and human cognition work together to entrench beliefs. People are more likely to notice and remember information that aligns with their existing attitudes and beliefs.¹³¹ Platforms, tuned to maximise engagement, learn those preferences and supply more of the same.¹³² The legal system then intervenes relatively late, asking jurors to adopt an attitude of disciplined neutrality for the duration of a single trial. Directions assume that jurors can temporarily set aside the broader narratives they have been exposed to about crime, youth offending, or lenient courts.¹³³ The empirical literature suggests that this is a demanding task, particularly where publicity has framed an accused as the embodiment of a broader social problem.¹³⁴

High Court decisions such as *Glennon* and *Dupas* rest on an explicit confidence that jurors will generally do as they are told: that they can put aside prejudicial publicity when directed, that they will not make their own inquiries, and that any residual risk is tolerable considering the importance of open justice and freedom of discussion.¹³⁵ That

¹²⁷ News & Media Research Centre (n 80) 8–10.

¹²⁸ ReachOut Australia, Beyond Blue and Black Dog Institute, *Inquiry into Social Media Impacts on Australian Society* (Submission to Joint Select Committee on Social Media and Australian Society, July 2024) 8.

¹²⁹ *Milli et al* (n 4) 1–2.

¹³⁰ Lisa K Fazio, David G Rand and Gordon Pennycook, ‘Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements’ (2019) 26 *Psychonomic Bulletin & Review* 1705.

¹³¹ Charles G Lord, Lee Ross and Mark R Lepper, ‘Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence’ (1979) 37(11) *Journal of Personality and Social Psychology* 2098.

¹³² Australian Competition and Consumer Commission (n 103) 63–5.

¹³³ New South Wales Law Reform Commission, *Jury Directions* (Report No 136, 2012) 25–7, 112–13.

¹³⁴ Nancy M Steblay *et al* (n 13).

¹³⁵ *Glennon* (n 7); *Dupas* (n 41).

confidence is partly normative – a statement about how jurors ought to behave – and partly empirical – a claim about how they in fact act. The research sketched above does not require a complete rejection of that stance. It does, however, weaken its factual foundation. It suggests that exposure to pre-trial publicity is more pervasive than the case law assumes, that directions do not reliably neutralise its effects, and that digital platforms actively reinforce the kinds of emotionally charged narratives most likely to prejudice an accused.¹³⁶ Chesterman, Chan, and Hampton's findings provided courts with some empirical support for trusting directions in a broadcast-media environment.¹³⁷ However, later work on social media and juror conduct shows that this support is incomplete unless it is extended to account for algorithmic curation and constant connectivity.¹³⁸ Burd and Horan, therefore, stand as an early doctrinally engaged warning that the 'trust the jury' assumption must be tested against evidence about how jurors encounter and use online information, rather than being treated as an article of institutional faith.¹³⁹

For doctrine, the implication is not that courts must abandon trials whenever a case has attracted online attention. A more realistic conclusion is that assessments of risk should be informed by a clearer understanding of how people process information in an algorithmic environment. As digital ecosystems evolve, maintaining the legitimacy of jury trial may require closer alignment between doctrinal assumptions about juror behaviour and what is known about juror psychology and online publicity.

D *Adequacy of Existing Safeguards – Doctrinal Evaluation*

Sub judice contempt remains a strong signal that publications creating a real and definite tendency to prejudice a trial are unlawful.¹⁴⁰ It discourages mainstream outlets from publishing the most inflammatory material and provides a route to sanction apparent breaches. In an algorithmic environment, however, its limits are more clearly exposed. Contempt is organised around identifiable acts of publication by specific actors.¹⁴¹ Viral content on social media is produced, modified and recirculated by countless users. Pursuing contempt proceedings against individual posters raises problems of scale and does little to

¹³⁶ Lori A Hoetger *et al* (n 10) 132-3.

¹³⁷ Michael Chesterman *et al* (n 37).

¹³⁸ Das *et al* (n 12) 230-3.

¹³⁹ Burd and Horan (n 121).

¹⁴⁰ *Attorney-General v John Fairfax & Sons Ltd* [1980] 1 NSWLR 362, 366–7 (Street CJ).

¹⁴¹ *Ibid.*

stem the derivative posts and comment threads that algorithms have already pushed through feeds.¹⁴²

The stay jurisdiction and related procedural levers are also under pressure.¹⁴³ In *Glennon* and *Dupas*, the High Court insisted that a permanent stay is justified only where no combination of directions and safeguards could realistically secure a fair trial.¹⁴⁴ That standard was framed with episodic, case-focused publicity in mind – a burst of reporting around an arrest or first trial. In the digital environment described earlier, the more significant risk is often long-running background exposure. A juror who has repeatedly seen a case framed as part of a broader story about youth crime or systemic failure may have internalised that narrative, even if they no longer recall specific details.¹⁴⁵ Traditional stay applications, built around particular articles or broadcasts that can be tendered and debated, struggle to capture that kind of diffuse conditioning.

Jury directions remain the cornerstone of the current framework.¹⁴⁶ Clear instructions about the presumption of innocence, the burden of proof and the need to decide only on the evidence provide an important normative anchor. Directions not to conduct online research or to discuss the case on social media are now routine.¹⁴⁷ The empirical literature, however, suggests that such directions have limited reach when attitudes have already formed or when information-seeking habits are deeply ingrained.¹⁴⁸ Telling a juror, after years of using a phone to verify facts and fill gaps, that they must rely solely on what is said in court requires a sharp departure from ordinary practice. Behavioural evidence suggests that some jurors will manage this in some matters, but not consistently.¹⁴⁹

Sequestration, device restrictions and judge-alone trials are sometimes presented as stronger answers. Isolating jurors and removing access to phones and the internet during the trial period should, in theory, reduce opportunities for mid-trial contamination. However, sequestration is used sparingly because it is expensive, disruptive, and onerous, and it does nothing to address pre-trial exposure.¹⁵⁰ Judge-

142 NSWLRC, *Contempt by Publication* (n 47).

143 *Ibid.*

144 *Glennon* (n 7); *Dupas* (n 41).

145 Raymond S Nickerson, ‘Confirmation Bias: A Ubiquitous Phenomenon in Many Guises’ (1998) 2(2) *Review of General Psychology* 175, 177–9.

146 Judicial Commission of New South Wales, *Criminal Trial Courts Bench Book* (online at 25 October 2025) ‘The Jury – Adverse Publicity in Media and on the Internet’ [1-450].

147 *Ibid* [1-490].

148 Nancy M Steblay *et al* (n 13); Tarika Daftary-Kapur *et al* (n 115).

149 Michael Chesterman *et al* (n 37); *TLRI* (n 110); Burd and Horan (n 121).

150 *TLRI* (n 110) 83-4, 96-8; NSWLRC, *Contempt by Publication* (n 47) 146-7.

alone trials sidestep concerns about lay susceptibility by relying on a professional decision-maker assumed to be better able to compartmentalise irrelevant material. Judges, however, live in the same information environment as everyone else and may also have been exposed to sustained coverage of controversial incidents. The concern is not that judges will deliberately disregard their oaths, but that background impressions formed through repeated digital exposure may operate below conscious awareness.¹⁵¹ Regular resort to the judge-alone procedure in response to publicity would also have legitimacy costs, sidelining community participation in the very cases that attract the most attention and that call for the application of ‘ordinary person’ standards. Judge-alone trials are, at best, an imperfect safeguard rather than a complete answer to digital-era prejudice.

On appeal, an accused can argue that pre-trial or mid-trial publicity, juror misconduct or mismanaged directions produced a miscarriage of justice.¹⁵² In practice, that safeguard is constrained by the structure of appellate review. Jury secrecy rules limit inquiry into what jurors saw or discussed, so appellate courts rarely have more than fragmentary material about extra-curial exposure.¹⁵³ Complaints about prejudicial publicity are assessed against the trial record, filtered through the directions the judge gave and the evidence that was formally tendered, and then subjected to a high threshold under proviso doctrines that ask whether any error caused a substantial miscarriage of justice.¹⁵⁴ The same assumptions that shape stay applications reappear: courts presume that clear directions were followed, that jurors did not conduct undisclosed online research, and that any residual risk fell short of undermining the verdict.¹⁵⁵ Cases in which convictions are set aside solely because of publicity or juror internet use remain exceptional.¹⁵⁶ For digital-era forms of exposure – such as cumulative encounters in feeds, algorithmically curated narratives, and AI-generated material glimpsed long before empanelment – it will often be impossible to reconstruct the influence in a manner that meets appellate standards of proof. Appeals, therefore, tend to reproduce systemic optimism about existing safeguards rather than providing a robust corrective when algorithmic publicity has shaped trial conditions.

¹⁵¹ Mahzarin R Banaji and Anthony G Greenwald, *Blindspot: Hidden Biases of Good People* (Bantam, 2013) 55–9.

¹⁵² Joel D Lieberman and Jamie Arndt, ‘Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence’ (2000) 6 *Psychology, Public Policy, and Law* 677.

¹⁵³ *Smith v Western Australia* [2014] HCA 3; (2014) 250 CLR 473, 483–4 [12].

¹⁵⁴ *Weiss v The Queen* [2005] HCA 81; (2005) 224 CLR 300, 313–15 [39]–[45]; *Criminal Code* s 668E(1A).

¹⁵⁵ *Glennon* (n 7); *Dupas* (n 41).

¹⁵⁶ *McEwen and Eldridge* (n 112) 111–13.

Taken together, the safeguards provide only a *prima facie* assurance of fairness and are heavily reliant on optimistic assumptions about human behaviour and institutional capacity. They presume that courts can identify and assess the relevant publicity, that jurors will comply with instructions that contradict entrenched digital habits, and that attitudes formed through years of curated feeds can be set aside for the duration of a trial. The law does not yet have a settled way to account for cumulative low-level exposure to prejudicial narratives. Nor has it grappled adequately with the role of algorithms in amplifying harmful framings. The difficulty of detecting – let alone remedying – online research that occurs before a juror ever enters a courtroom remains largely unaddressed.¹⁵⁷ Nor has it developed a coherent view about the responsibilities, if any, that should rest on platforms whose systems help shape those narratives. The following sections address these questions by examining, first, regulatory measures directed at digital platforms' algorithmic publicity in other jurisdictions and, secondly, the potential for court-supervised AI monitoring tools to provide judges with a clearer view of the digital terrain surrounding the cases they are asked to try.

E *Comparative Regulatory Responses*

Building on that diagnosis, this section argues that comparative experience marks out the outer boundaries of plausible responses: UK and EU regimes show how narrowly drawn, systemic duties on platforms and high-risk AI can support fair-trial protections; US and Chinese approaches illustrate opposite pathologies of under- and over-regulation; and smaller jurisdictions such as Estonia model assistive, court-centred uses of AI.

In the UK, the *Online Safety Act 2023* (UK) ('OSA') establishes a statutory duty of care for major user-to-user and search services, primarily implemented through systemic risk assessments of how their recommender and search systems surface and disseminate illegal or harmful content.¹⁵⁸ Providers must identify the patterns through which their ranking tools disseminate such material to users and take reasonable steps to mitigate those risks.¹⁵⁹ OSA does not refer to *sub judice* contempt or juries, but its focus on system-level risks is directly relevant to the problem identified earlier.¹⁶⁰ Once a potential juror or other user searches for a case, engagement-based algorithms may present them with a stream of content asserting guilt or framing the

¹⁵⁷ Ibid.

¹⁵⁸ *Online Safety Act 2023* (UK) ('OSA') pt 3 chs 2–3.

¹⁵⁹ Ibid.

¹⁶⁰ Ibid.

accused as emblematic of wider social failure, a pattern documented in empirical work on juror internet use and ‘incidental’ exposure.¹⁶¹ For Australia, the lesson is that Parliament could require large platforms to assess whether their systems are amplifying clearly prejudicial material about pending proceedings, and to address that amplification at a design level, while leaving ordinary reporting and criticism untouched, provided any duty is tightly confined.

The EU’s *Artificial Intelligence Act 2024* (EU) (*‘AI Act’*) takes a risk-based approach.¹⁶² It classifies specific applications, including those tied to justice systems, as ‘high-risk’ and subjects them to documentation, pre-deployment testing, human oversight, and monitoring.¹⁶³ Many AI tools are also bound by transparency obligations, such as informing users when they interact with a chatbot or when content has been artificially generated or altered. Member States are moving towards labelling regimes for deepfakes that are likely to mislead the public.¹⁶⁴ Two features of this framework are especially relevant. First, the insistence on labelling synthetic media offers a clear analogue for a requirement that AI-generated or heavily manipulated images or videos depicting accused persons be identified as such while proceedings are ongoing.¹⁶⁵ Secondly, the idea that high-risk systems must be auditable and under human control maps onto the way any court-supervised publicity-monitoring tool should be treated in Australia – as a judicial support system whose methods and outputs can be scrutinised, rather than a black box feeding quietly into decisions.¹⁶⁶

Estonia, often invoked as experimenting with ‘AI judges’, offers a more modest and practically instructive example. Courts there have invested in digital infrastructure and utilise automation for tasks such as transcription, anonymisation of judgments, and routing straightforward payment-order procedures.¹⁶⁷ At the same time, final decisions remain with human judges.¹⁶⁸ The government has rejected replacing judicial decision-making with algorithms.¹⁶⁹ Estonia,

¹⁶¹ Das *et al* (n 12) 230-3.

¹⁶² Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (*Artificial Intelligence Act*) (*‘AI Act’*).

¹⁶³ Ibid arts 9-15.

¹⁶⁴ Ibid art 52.

¹⁶⁵ Ibid.

¹⁶⁶ Ibid arts 16-23.

¹⁶⁷ Oxford Institute of Technology and Justice, ‘Estonia’ (Web Page) <<https://www.techandjustice.bsg.ox.ac.uk/research/estonia>>.

¹⁶⁸ Ibid.

¹⁶⁹ Republic of Estonia Ministry of Justice and Digital Affairs, ‘Estonia Does Not Develop AI Judge’ (Web Page, 16 February 2022) <<https://www.justdigi.ee/en/news/estonia-does-not-develop-ai-judge>>.

therefore, stands as a concrete model of AI used as administrative support rather than adjudicator: automation is embedded in information-management functions, while legal judgment stays human. That configuration sits close to the kind of court-supervised monitoring tool proposed later in this article, in which AI organises and surfaces public information but does not determine outcomes.

The US occupies a distinct position due to its First Amendment tradition, which leaves little room for legislation imposing platform duties centred on fair trials.¹⁷⁰ In *Sheppard v Maxwell*, the Supreme Court quashed a murder conviction due to intense and uncontrolled publicity, emphasising that trial judges have a positive duty to manage media coverage through measures such as venue changes and sequestration, rather than relying on statutory limits on reporting.¹⁷¹ That model has since been tested by live-streamed proceedings and pervasive social media commentary, exemplified by the *Depp v Heard* defamation trial, where the proceedings themselves became online content.¹⁷² For Australia, the American experience highlights the limitations of relying almost entirely on case-by-case trial management when platform architecture and recommendation systems that generate prejudicial narratives remain unaddressed.

China sits at the opposite end of the regulatory spectrum. Douyin, the domestic counterpart to TikTok, operates within a tightly controlled system of real-name registration, extensive moderation, and data localisation obligations.¹⁷³ Those levers make it straightforward for authorities to suppress content regarded as harmful to social order, including material that might prejudice trials or criticise courts, and to steer crime narratives in line with official priorities.¹⁷⁴ From a fair-trial perspective, Douyin therefore marks the other boundary of the policy space: prejudicial publicity can be curtailed by digital architecture, but only by sacrificing anonymity, independent media and open criticism that are central to the legitimacy of criminal adjudication in a liberal democracy.

Other examples confirm that media ecosystems and procedural choices interact rather than supplying templates to copy. Canadian courts have been more willing to order judge-alone trials in highly

¹⁷⁰ Evelyn Mary Aswad, 'The Future of Freedom of Expression Online' (2018) 17 *Duke Law & Technology Review* 26.

¹⁷¹ *Sheppard v Maxwell*, 384 US 333, 349–52, 358–63 (1966).

¹⁷² *John C Depp II v Amber Laura Heard* (Fairfax County Cir Ct (Va), No CL-2019-2911, 1 June 2022).

¹⁷³ The WeChat Agency, 'What is Douyin, the Chinese TikTok Equivalent?' (Web Page, 18 November 2025) < <https://www.thewechatagency.com/what-is-douyin/> >.

¹⁷⁴ US–China Economic and Security Review Commission, *Censorship Practices of the People's Republic of China* (Report, 20 February 2024) 13–17.

sensitive matters, limiting direct exposure of lay jurors to publicity.¹⁷⁵ The contrasting outcomes of *Depp v News Group Newspapers Ltd* in London (a judge-alone libel trial) and Depp's later jury trial in Virginia illustrate how the same allegations can play out differently where a single judge decides liability in a contained media environment and, by contrast, a jury deliberates under intense online scrutiny.¹⁷⁶

Overall, the comparative material outlines boundaries rather than providing a blueprint. The *OSA* and the *AI Act* demonstrate that legislatures can impose risk-based obligations on platforms and AI providers, while also requiring transparency about how algorithms operate, without compromising the principles of open justice.¹⁷⁷ The US highlights the limits of leaving the burden with individual trial judges in a rapidly changing digital environment.¹⁷⁸ The Douyin model signals what should be avoided if fair-trial protections are to remain compatible with liberal commitments. Within that space, a plausible course for Australia is modest. It involves three elements: narrowly targeted duties on major platforms in relation to clearly prejudicial content about pending proceedings; transparency and labelling requirements for synthetic media depicting accused persons; and the careful use of court-supervised AI tools that support, rather than displace, existing doctrines.

F *Court-Supervised AI as a Targeted Response*

If courts are to take digital publicity seriously, they need a clearer view of what is actually circulating about a case. Currently, counsel presents a few headlines or screenshots, and judges are asked to reconstruct the environment from them.¹⁷⁹ A more systematic approach would be a court-supervised AI tool that scans defined public sources – including major news sites, public social media posts, and search engine results – for content linked to a proceeding. Using basic language processing, it could group material into broad categories, distinguish between neutral reporting and commentary, and identify posts that present guilt as settled.

The tool would not make findings on prejudice; its role is to provide courts with a more grounded sense of online attention than *ad hoc* exhibits can offer. A judge could see that attention to a case is intensifying, that posts asserting guilt are being widely reshared, or that

¹⁷⁵ *R v Ghomeshi* (2016) ONCJ 155.

¹⁷⁶ *Depp v News Group Newspapers Ltd* [2020] EWHC 2911 (QB) ('*Depp v News*').

¹⁷⁷ *OSA* (n 158); *AI Act* (n 162).

¹⁷⁸ *Depp v News* (n 176).

¹⁷⁹ Amy Elvidge, 'Trying Times: The Right to a Fair Trial in the Changing Media Environment' (LLB(Hons) Dissertation, University of Otago, 2008) 52.

a synthetic clip purporting to show the offence has begun to circulate. That picture could then inform decisions on adjournment, the timing and content of directions, individual juror questioning or, in an extreme case, a change of venue or a judge-alone order.

The primary speech risk is chilling: if users believe that ordinary discussion of criminal matters contributes to a form of judicial surveillance machine, they may self-censor far beyond what contempt law requires.¹⁸⁰ That risk can be reduced by keeping the remit narrow – confined to clearly case-linked content during a defined pre-trial and trial window – and by publishing clear guidance that the tool exists to inform fair-trial decisions, not to generate targets for prosecution or disciplinary action.¹⁸¹

The tool should operate only on material that any member of the public could lawfully view; it should not search by juror names or track individual users. Any report that informs a judicial decision in the case should be available to both parties.

Those baseline constraints – a narrow remit and public-only sources, with shared access to outputs – should be built into the system. The aim is not to create a new investigative arm of the state, but to provide a limited, transparent support tool that helps courts see the publicity environment they are already required to manage.

The same system could also have a limited appellate role. Suppose courts retained an archived record of the publicity reports generated during the life of a proceeding. In that case, an appeal court considering whether a verdict was unreasonable on the evidence, or whether there has been a miscarriage of justice, could weigh the trial record against a contemporaneous picture of the surrounding publicity rather than relying on fragmentary affidavits or recollections. The tool would not alter the legal tests on appeal, but it could sharpen the factual premise on which those tests are applied.

Seen in that light, court-supervised AI is a support for existing doctrine, not a new source of legal norms. It does not alter the elements of contempt, the principles governing stays or the content of jury directions, and it cannot cure deeper problems of digital echo chambers and long-term attitude formation. What it can do is provide courts with a more concrete empirical basis for applying those doctrines in an environment where online publicity is prevalent. With a narrow mandate and firm safeguards, a monitoring system of this kind could

¹⁸⁰ Moritz Büchi, Noemi Festic and Michael Latzer, 'The Chilling Effects of Digital Dataveillance: A Theoretical Model and an Empirical Research Agenda' (2022) 9(1) *Big Data & Society* 1.

¹⁸¹ International Commission of Jurists, *Digital Technologies and Human Rights: A Legal Framework* (Briefing Paper, May 2022) 9 <<https://www.icj.org/wp-content/uploads/2022/05/Digital-Technologies-and-Human-Rights-Briefing-Paper-FINAL-VERSION-May-2022.pdf>>.

help align legal decision-making with the realities of algorithmic publicity while preserving the primacy of human judgment and the core commitments of the fair-trial guarantee.

IV Conclusion and Recommendations

The analysis in this article suggests that social media and algorithmic publicity do not make fair criminal jury trials impossible. They do, however, make the current doctrinal settlement more contingent than the leading authorities acknowledge. A framework built around episodic broadcast publicity and identifiable publishers, underpinned by intuitive confidence in juries, now operates in an environment of continuous, personalised feeds, synthetic media and AI-generated narratives.¹⁸² As that gap widens, the risk is not only that individual trials may be compromised, but also that public confidence in jury verdicts will erode, particularly among younger, digitally native cohorts whose understanding of the criminal justice system is shaped primarily online.¹⁸³

Doctrinally, Australian law remains structured and principled. *Hinch* anchors *sub judice* contempt in the requirement of a real and definite tendency to prejudice justice, signalling that egregious distortions of pending proceedings remain unlawful.¹⁸⁴ *Glennon* and *Dupas* affirm that permanent stays are exceptional and that, in most cases, fairness can be preserved through directions, juror questioning and procedural safeguards.¹⁸⁵ *Evidence Act* discretions provide a check on undue prejudice within the trial record, and law reform commissions have recognised that contempt and suppression doctrines strain under digital conditions, even as basic assumptions about the manageability of publicity and jury reliability remain intact.¹⁸⁶

Set against this is the empirical and technological picture. Potential jurors live in media environments configured by engagement-optimised algorithms.¹⁸⁷ Emotive or accusatory content is preferentially surfaced and repeated, deepfakes and other synthetic media blur the line between authentic evidence and fabrication, and generative AI condenses

¹⁸² Haochen Sun, 'Regulating Algorithmic Disinformation' (2023) 46(4) *Columbia Journal of Law & the Arts* 367, 374–5, 377–9.

¹⁸³ Monica K Miller *et al*, 'Trust in the Jury System: A Comparison of Australian and US Samples' (2021) 28(6) *Psychiatry, Psychology and Law* 823.

¹⁸⁴ *Hinch* (n 25).

¹⁸⁵ *Glennon* (n 7); *Dupas* (n 41).

¹⁸⁶ *Evidence Act* (n 69).

¹⁸⁷ Itxaso Dominguez de Olazábal and Jan Penfrat, *A Rights-Based Digital Fairness Act: Background Paper on Protecting People from Manipulative Design and Data Exploitation* (European Digital Rights (EDRi), October 2025) 33, 38, 46.

available material into apparently authoritative narratives.¹⁸⁸ Research on juror behaviour indicates that pre-trial publicity can influence verdict preferences, that directions only partly mitigate these effects, and that people tend to seek and remember information that confirms their existing attitudes.¹⁸⁹ For many users, especially younger cohorts, searching for a name or case online is a routine habit rather than a deliberate act of defiance.¹⁹⁰

Comparative experience illustrates both the possibilities and the risks associated with regulatory intervention. Risk-based duties on platforms and AI providers, as developed in EU and the UK, show that legislatures can require systemic attention to how recommender and search systems surface harmful content and to the governance of high-risk applications.¹⁹¹ By contrast, models that leave almost all responsibility with courts in a strong free-speech culture expose the limits of case-by-case trial management.¹⁹² At the same time, heavily controlled information environments demonstrate that rapid suppression of prejudicial content is achievable only at a serious cost to freedom of expression and independent scrutiny.¹⁹³ More modest experiments in using AI to support, rather than displace, judicial decision-making suggest a better path: technology that assists courts in seeing and managing the publicity environment, without turning platforms into censors or judges into supervisors of online speech.¹⁹⁴

Social media and AI-driven publicity undermine the right to a fair trial, where they create durable prejudicial narratives that existing tools are poorly equipped to detect or neutralise.¹⁹⁵ Existing doctrines assume episodic exposure, traceable sources, and influence that can be mitigated through corrective directions.¹⁹⁶ The digital environment provides cumulative exposure to diffuse origins and attitudes that are resistant to simple instruction. The response lies not in abandoning jury trials or open justice, but in recalibrating doctrine, procedure, and

¹⁸⁸ William J Brady *et al* (n 84).

¹⁸⁹ Amy L Otto, Steven D Penrod and Hedy R Dexter, 'The Biasing Impact of Pretrial Publicity on Juror Judgments' (1994) 18(4) *Law and Human Behavior* 453.

¹⁹⁰ Emily Baker, 'Jurors Using Social Media to Look Up Victims and Criminals, Researcher Says' *ABC News* (online, 21 August 2019) <<https://www.abc.net.au/news/2019-08-21/jurors-using-social-media-to-look-up-victims-and-criminals/11435874>>.

¹⁹¹ *OSA* (n 158); *AI Act* (n 162).

¹⁹² Gavin Phillipson, 'Trial by Media: The Betrayal of the First Amendment's Purpose' (2008) 71(4) *Law and Contemporary Problems* 15.

¹⁹³ Human Rights Watch, 'World Report 2025: China' (Web Page, Human Rights Watch, 2025) 'Freedom of Expression' <<https://www.hrw.org/world-report/2025/country-chapters/china>>.

¹⁹⁴ Australasian Institute of Judicial Administration, *AI Decision-Making and the Courts: A Guide for Judges, Tribunal Members and Court Administrators* (Report, 2023) 9–11, 59–60.

¹⁹⁵ VLRC, *Contempt of Court: Report* (n 2).

¹⁹⁶ *Glennon* (n 7); *Hinch* (n 25); *Dupas* (n 41).

regulation so that fair-trial principles are applied with a realistic understanding of how information now circulates. The remainder of this section sets out three recommendations.

1 *Doctrinal and Procedural Refinement*

Courts should make more explicit use of contemporary knowledge about digital publicity when applying existing tests. Applications for stays and venue changes should include structured consideration of the scale and persistence of all forms of publicity, including online coverage, rather than relying on a few prominent stories. Jury directions should be updated to address social media and AI in concrete terms, including explanations of how algorithms reinforce prior interests and warnings about deepfakes and AI-generated commentary. Appellate courts should exercise caution when relying solely on intuitive assertions of juror resilience, particularly in light of empirical work that suggests otherwise.¹⁹⁷ Guidance on judge-alone trials should acknowledge both the protective potential and the legitimacy costs of sidelining community participation in clear cases.

2 *Targeted Platform Regulation*

There is scope for narrowly drawn statutory duties on major platforms that address fair-trial risks directly, without collapsing into broad censorship. One option is a court-triggered mechanism. Notification of serious pending criminal proceedings would activate a limited duty of care on designated platforms to assess how their recommender and search systems surface case-linked content and to reduce algorithmic promotion of material that asserts guilt as fact or circulates synthetic media purporting to depict the alleged offence. Another approach is to require clear labelling of AI-generated or heavily manipulated images, audio, or video that refer to identifiable accused persons while proceedings are ongoing, drawing on emerging EU deepfake labelling models.¹⁹⁸ Any such scheme must be tightly bounded in subject matter and duration, framed at system-design level rather than *ad hoc* takedowns, and subject to independent oversight to avoid gradual expansion into general speech control.

3 *Court-Supervised AI and Institutional Capacity*

As suggested above, a narrow, court-supervised publicity-monitoring tool, confined to public content and subject to clear safeguards, could provide judges with a more accurate view of the online environment

¹⁹⁷ Lori A Hoetger *et al* (n 10).

¹⁹⁸ *AI Act* (n 162).

surrounding a case and inform decisions about adjournment, directions, and jury management. To be institutionally credible, any such system should be treated as a high-risk judicial support application: its purpose must be tightly defined, its data sources and classification methods documented, its operation transparent and auditable, and its outputs available to both parties, with decision-making authority remaining squarely with the judge.

V Objections and Limits

Any package of reforms of this kind is likely to attract predictable objections. One is that the problem is overstated and that juries continue to cope well with publicity. This article does not deny that many trials remain fair in practice; the claim is narrower: that the gap between doctrinal assumptions and the realities of algorithmic publicity is widening, and that calibrating doctrine and procedure to that reality strengthens, rather than weakens, the jury system. A second concern is that platform duties and court-supervised monitoring tools risk becoming a back-door form of speech control or digital surveillance.¹⁹⁹ The model proposed here is deliberately modest: duties are confined to the algorithmic promotion of clearly prejudicial content during a defined pre-trial and trial window; monitoring tools operate only on publicly available material; and neither creates new categories of prohibited speech. Even so, these measures can only mitigate, not eliminate, the effects of long-running algorithmic echo chambers.

VI Concluding Reflections

These recommendations are incremental rather than revolutionary. They assume that the presumption of innocence, the fair trial guarantee and open justice remain non-negotiable features of the Australian criminal process.²⁰⁰ What changes is the set of institutional and regulatory supports that make those commitments credible for jurors whose world is mediated by algorithms. Safeguarding justice and dignity for digital-native jurors will require the legal system to acknowledge how contemporary publicity operates in practice and to adjust its tools so that criminal adjudication remains both fair in substance and seen as fair by a public whose perceptions are increasingly shaped online. A natural next step is empirical work that tracks how potential jurors actually encounter algorithmically curated

¹⁹⁹ Ian Barber and Maria Paz Canales, 'What Would a Human Rights-Based Approach to Platform Regulation Look Like?' (Blog Post, Global Partners Digital, 30 July 2024) <<https://www.gp-digital.org/what-would-a-human-rights-based-approach-to-platform-regulation-look-like/>>.

²⁰⁰ *Dietrich* (n 9); *Hogan v Hinch* (2011) 243 CLR 506, 530 [20] (French CJ).

content about criminal cases, and doctrinal engagement with future Australian decisions that grapple explicitly with algorithmic publicity, so that law and practice can be calibrated against observed – rather than assumed – patterns of exposure.