UNE Kirby Seminar, 5 September 2024. School of Law, University of New England

# DISTORTIONS OF LEGAL AND OTHER RELATIONSHIPS: HAS THE INTERNET CHANGED OUR PERCEPTIONS OF REALITY AND EACH OTHER?

The Hon. Steven Rares KC*

*The Hon. Michael Kirby AC CMG*

It is an honour to be asked to give this seminar. I first heard of the Hon. Michael Kirby AC CMG when I was a law student and working as a research assistant for a member of the Parliament. Michael was then the inaugural chairman of the Australian Law Reform Commission. He wrote frequent personal letters to my boss enclosing copies of each new discussion paper and report that the Commission produced. Each was of high quality, reflecting the boundless intellectual energy and depth of thought that Michael continued to bring to each of the many responsibilities he took on, initially for this nation and later for the benefit of all the world's peoples.

My next encounters with him came following my call to the bar when I became a member of his old floor of 12 Wentworth Chambers in Sydney, and I met him at floor dinners and functions. Our old clerk, Greg Isaac, would extol Michael's ability to turn any brief to advise around in 24 hours. Soon after, in 1984, he became a very young President of the Court of Appeal of the Supreme Court of New South Wales. Within a year he had transformed the culture of that Court from one where senior as well as junior counsel, felt that they were blood sport for the bench, to one in which the Court was interested in what counsel had to say and politely engaged in the hearings. Michael travelled and presented papers relentlessly while also fully performing his judicial duties, writing thoughtful and, typically, thorough judgments that often developed the law. He took on other roles, including as patron of the Friends of the Museum of Applied Arts and Sciences, of which my late father was then president. They became close friends, and Michael often reminds me of this. As his colleague on the Court of Appeal, the late Hon. Simon Sheller AO QC wrote for Michael's biography in the Oxford Companion to the High Court of Australia, he listed his hobby in

*Who's Who* as *"work"*.

In 1996 Michael was appointed a justice of the High Court of Australia, on which he served with distinction until 2009. His judgments, often through his dissents, like Lord Denning's in my student days, were for law students, inspirational in seeking to show how the general law can adapt to meet the enormous social changes that had occurred over our lifetimes. As is well known, Michael became a trailblazer against all forms of discrimination, including when he came out early in his time on the High Court.

I would not be able to give this paper if I listed and expanded on all **of** Michael's achievements and contributions, including to the Human Genome Project and as a Special

* Formerly a Judge of the Federal Court of Australia from 2006 to 2023 and currently an arbitrator, mediator and Adjunct Professor at the University of New South Wales Faculty of Law and Justice. This paper is adapted from a speech given on 5 September 2024 at the University of New England as part of the 2024 Kirby Seminar Series. The author thanks Scarlett de Vine, his research assistant, for her help in the preparation of this paper. Any errors are the author's alone.

Representative of the Secretary-General of the United Nations for Human Rights in Cambodia. The University of New England has rightly honoured a great Australian in naming this seminar series after Michael Kirby.

**Introduction**

In what follows I want to reflect on how various societies have grappled with the concepts of truth, lies and propaganda in human communication. The modern age has thrown up new challenges to our ability to discern what is objectively true, or sufficiently certain, from what is not. We like to think that the enforcement of rights and obligations under the law follows from findings of fact by judges and juries of where the truth lies. But, the different standards of proof in civil and criminal matters show that even the law does not proceed on the basis that there can only be absoluteness about the correctness of any finding of fact.

Throughout history, humans have had to grapple with whether something is real or a deception, and what are the consequences. The monotheistic religions of Judaism, Christianity and Islam all reflect on the biblical story of Adam and Eve. This tells us that God forbade Adam from eating the fruit of the tree of the knowledge of good and evil in the garden of Eden[1], telling him *"for when you eat from it you will certainly die"*. The serpent deceived Eve by telling her that she will certainly not die and saying, *"For God knows that when you eat from it your eyes will be opened, and you will be like God, knowing good and evil"*. When God discovered that Adam and Eve had eaten the forbidden fruit he made them mortal. What the serpent said was, in one sense literally true. That is because the fruit itself was not poisonous and, so, not immediately fatal. But, in another sense, the serpent lied because the consequence of eating the forbidden fruit was that God did make certain that man would no longer be immortal.

In his famous speech in *Smith v Chadwick*[2], Lord Blackburn explained how a trick, like the serpent's, is a fraud saying:

> The defendants might honestly believe that the shares were a capital investment, and that they were doing the plaintiff a kindness by tricking him into buying them. I do not say this is proved, but if it were, if they did trick him into doing so, they are civilly responsible as for a deceit. And if with intent to lead the plaintiff to act upon it, they put forth a statement which they know may bear two meanings, one of which is false to their knowledge, and thereby the plaintiff putting that meaning on it is misled, I do not think they can escape by saying he ought to have put the other. If they palter with him in a double sense, it may be that they lie *like* truth; but I think they lie, and it is a fraud. Indeed, as a question of casuistry, I am inclined to think the fraud is aggravated by a shabby attempt to get the benefit of a fraud, without incurring the responsibility.

Although deception and fraud have a timeless quality about them, their existence also raises the philosophical question that Francis Bacon posed in 1625 in the opening sentence of his essay *Of Truth*[3], *"'What is truth'? said Jesting Pilate and would not stay for an answer"*. Sir Owen Dixon used that sentence to end a speech entitled *Jesting Pilate*[4].

Much ink has been spilt discussing the real purpose of Pilate's words and actions[5]. Various hypotheses have emerged. Some think that Pilate's question was a sarcastic retort to Jesus' assertion that *"I came into the world, that I should bear witness unto the truth"*. Others, that Pilate posed rhetorically a philosophical question about how 'truth' is defined.

---

[1] Genesis 2 and 3.
[2] (1884) 9 App Cas 187 at 201 (emphasis in original).
[3] *Essays or Counsels, Civil and Moral.*
[4] Owen Dixon, 'Jesting Pilate' in Woinarski (ed) *Jesting Pilate and Other Papers and Addressed by the Right Honourable Sir Owen Dixon* (Law Book Co, Sydney 1965) p 4.
[5] described in John 18.37-.40.

In his last address as Chief Justice of New South Wales, *Truth and the Law*[6], the Hon JJ Spigelman AC, questioned the correctness of both Bacon's and Sir Owen's description of Pilate's hasty exit. Chief Justice Spigelman explained that, in the biblical accounts, Pilate left the room briefly to consult, what his Honour characterised as the jury, before returning to the trial. The Chief Justice then examined in his usual scholarly manner, the way in which judicial proceedings in both common law and civil law jurisdictions engage in fact finding for the purpose of resolving a dispute.

Bacon's version of Pilate's rapid departure is reminiscent of the current age's impatience with deep thought or analysis of issues in everyday life. That impatience may be connected to casualness in the accurate reporting of facts in the ever-changing focuses of the now pervasive 24-hour news cycle. Rarely do journalists dwell on giving us the facts or examining major or minor events beyond recounting a superficial account that may or may not distort what happened or was said. And, often, the version presented to the reader, viewer or listener has been tailored by his or her internet or social media provider based on that individual's previous browsing history. This presents each person with a different 'truth' about the same event because it is reported substantively differently. Thus, the 'news' depends on what the internet or social media provider's algorithms discern is the viewpoint that will match the perceived beliefs or prejudices of each audience member.

As I will elaborate later, throughout history, different news sources report an event in ways that can cause each person's perception of what is true or accurate to vary. One perpetually available means of doing this is through propaganda.

Truth and justice have an interrelationship. Long before Pilate posed his question, Plato advocated a vision of justice based on an enforced totalitarian reality that Sir Karl Popper summarised as follows[7]:

> Plato draws his final conclusion that any changing or intermingling within the three classes must be injustice, and that the opposite, therefore, is justice: 'When each class in the city minds its own business, the money-earning class as well as the auxiliaries and the guardians, then this will be justice.' This conclusion is reaffirmed and summed up a little later: 'The city is just… if each of its three classes attends to its own work.' But this statement means that Plato identifies justice with the principles of class rule and class privilege. For the principle that every class should attend to its own business means, briefly and bluntly, that *the state is just if the ruler rules, if the worker works, and if the slave slaves.*

Both Chief Justice Spigelman in the above address[8] and, in 2019, Justice Stephen Gageler AC, in his address to Harvard Law School, *Alternative Facts in the Courts*[9] referred to another remark of Sir Owen Dixon in response to a woman at a dinner party who suggested how wonderful it was to dispense justice:

> I do not have anything to do with justice, madam. I sit on a court of appeal, where none of the facts are known. One third of the facts are excluded by normal frailty and memory; one third by the negligence of the profession; and the remaining third by the archaic laws of evidence.

## Truth in the court room

As all lawyers know, in the common law tradition there are two standards of proof, *first*, in a criminal case, proof beyond reasonable doubt, and *secondly*, in a civil proceeding, proof on the balance of probabilities. Stating this difference immediately throws up the stark realisation

---

[6] (2011) 85 ALJ 746 at 747.

[7] Karl R. Popper, *The Open Society and its Enemies*, *Volume 1, the Spell of Plato* (Routledge & Kegan Paul, 5th ed, 1966) at 90 (footnote references omitted, emphasis in original).

[8] (2011) 85 ALJ 746 at 746-747.

[9] (2019) 93 ALJ 585 at 588.

that curial fact finding by a jury, a judge or an appellate court, does not necessarily provide certainty about the truth of what happened on any given occasion.

Thus, something that a court finds happened, such as "A murdered B", and so can be said to be 'true' in a civil proceeding, may not be found 'true' on the same or similar evidence in a criminal proceeding because of the different applicable standard of proof. This is what happened in *Helton v Allen*[10]. There Mr Helton was tried twice for the murder of the testator. After the jury's guilty verdict was reversed on appeal, he was acquitted at the second trial. However, when Mr Helton sought to prove the will of the deceased under which he would take the greater part of her property, others challenged his entitlement to do so on the rule of public policy that a person who kills his or her benefactor cannot take under the benefactor's will or on his or her intestacy[11]. Mr Helton relied on his acquittal to get around this. Dixon, Evatt and McTiernan JJ retorted[12]:

> There is, however, no trace of any such conception in the history of the principle that by committing a crime no man could obtain a lawful benefit to himself. To qualify the rule in the manner suggested would, we think, amount to judicial legislation.

Thus, at a criminal trial, you may be innocent until proven guilty beyond reasonable doubt, but, in a civil trial, you can still be proven guilty of the same or a lesser included offence (e.g., manslaughter instead of murder). This is provided, of course, that at the civil trial the judge or jury makes the finding on the balance of probabilities, having regard to the nature of the subject matter of the proceeding and the cause of action or defence, as well as the gravity of the allegations[13].

**Truth beyond the courtroom**

Most importantly, our need for truth exists outside the courtroom. Each of us needs accurate information to navigate daily life. We process material that comes to us via various media, including our own five senses, in order to make almost every decision, from the everyday trite function of walking on a safe footing, to the most serious, sometimes life and death situations. In that processing, we necessarily make judgments as to what is sufficiently reliable so that we can act on it. Like the serpent in the Garden of Eden, there are many sources of information capable of giving us a false sense of certainty or security. Since the internet and its progeny of social media have become part of our habitual environment, the potential readily accessible sources of information and misinformation that we perceive and may seek to act on have multiplied beyond our ancestors' imaginations.

Sir Karl Popper was an adherent of the absolutist theory of truth, which he traced back to Aristotle. He said that a statement is true if and only if it agrees with the facts it describes[14]. Popper explained the interconnection between propaganda that disseminates lies with the acquisition and maintenance of political and societal power in the following passage[15]:

> If we consider Plato's blunt admission that his Myth of Blood and Soil is a propaganda lie, then the attitude of the commentators towards the Myth is somewhat puzzling. Adam, for instance, writes: 'without it, the present sketch of a state would be incomplete. We require some guarantee for the permanence of the city...; and nothing could be more in keeping with the *prevailing moral and religious*

[10] (1940) 63 CLR 691.
[11] Ibid at 709.
[12] Ibid at 710.
[13] See *Evidence Act 1995* (Cth) s 140(2) and its analogues, reflecting the principles in *Bringinshaw v Bringinshaw* (1938) 60 CLR 336 at 362 per Dixon J.
[14] Karl R. Popper, *The Open Society and its Enemies*, *Volume 1, the Spell of Plato* (Routledge & Kegan Paul, 5th ed, 1966) at 273 n 23.
[15] Ibid (citations omitted, emphasis in original).

> *spirit* of Plato's... education than that he should find that guarantee in *faith rather than in reason.* I agree (though this is not quite what Adam meant) that nothing is more in keeping with Plato's totalitarian morality than his advocacy of propaganda lies. But I do not quite understand how the religious and idealistic commentator can declare, by implication, that religion and faith are on the level of an opportunist lie. As a matter of fact, Adam's comment is reminiscent of Hobbes' conventionalism, of the view that the tenets of religion, although not true, are a most expedient and indispensable political device. And this consideration shows us that Plato, after all, was more of a conventionalist than one might think. He does not even stop short of establishing a religious faith 'by convention' (we must credit him with the frankness of his admission that it is only a fabrication), while the reputed conventionalist Protagoras at least believed that the laws, which are our marking, are made with the help of divine inspiration.

## Getting to the truth

The First Amendment to the Constitution of the United States of America, among others, prohibits Congress from making any law *"abridging the freedom of speech or of the press"*. The aspiration behind this restriction is that by permitting all persons to say or opine whatever they wish, society will benefit by being able to discern the truth from what is said. Lord Eldon LC captured the idea, when explaining a barrister's role in the administration of justice (albeit, subject to the ethical constraints of his or her profession). The Lord Chancellor said: *"truth is best discovered by powerful statements on both sides of the question"*[16]. But unlike Anglo-Australian ethical constraints on lawyers in their dealings with courts and the public, the First Amendment ensures that in the United States, no such constraints can be imposed on anyone in their public expression of views. This includes the multinationals, colloquially known as 'Big Tech', which own and operate most of the structural internet and social media platforms that are now embedded in our ordinary lives, and facilitate how we often speak or communicate.

In Australia, the implied constitutional freedom of communication on government and political matter derives from the text and structure of the Commonwealth *Constitution*[17]. It limits all domestic legislative powers to make laws that burden communication to, or between, electors that are not reasonably and appropriately adapted to serve a legitimate end the fulfilment of which is compatible with the maintenance of our constitutionally prescribed system of representative government[18].

The free flow of information can convey material that has a relation to truth or objectivity. But, that relation can be pure as the driven snow or as putrid as a sewer. Ironically, according to a 2016 *Time* magazine article on Google[19], the famous aphorism attributed to Abraham Lincoln *"you can fool all of the people some of the time and some of the people all of time, but you cannot fool all of the people all of the time"* has no contemporary evidence linking it to Lincoln.

In 1931, Joseph Goebbels wrote in haunting language about how Nazism operated[20]:

> No other political movement has understood the art of propaganda as well as the National Socialists. From its beginnings, it has put heart and soul into propaganda. What distinguishes it from all other political parties is the ability to see into the soul of the people and to speak the language of the man in the street. It uses all the means of modern technology. Leaflets, handbills, posters, mass demonstrations,

---

[16] *Ex parte Lloyd* (1822) Mont 70 at 72n being a note to *Ex parte Elsee* (1832) Mont 69.
[17] *1900* (Cth).
[18] cf *Lange v Australian Broadcasting Corporation* (1997) 189 CLR 520 at 567-568 per Brennan CJ, Dawson, Deane, Toohey, Gaudron, McHugh and Gummow JJ; *Unions New South Wales v New South Wales* (2019) 264 CLR 595 at 607 [14] per Kiefel CJ, Bell and Keane JJ, at 621-623 [64]-[68] per Gageler J, at 640-641 [117]-[119] per Nettle J at 651-652 [155] per Edelman J, cf too at 641 [122] per Gordon J.
[19] David B. Parker, 'The Real Story Behind Abraham Lincoln's 'You Can Fool All the People' Quote', *Time* (Blog Post, 20 February 2016) <https://time.com/4231031/fool-all-the-people-lincoln-quote/>.
[20] Joseph Goebbels "Wille und Weg" W*ille und Weg* 1 (1931) at 2-5.

> the press, stage, film and radio — these are all tools of our propaganda. Whether or not they serve or harm the people depends on the use to which they are put.

Since first entering the field for the 2016 Republican Party nomination for election as President, Donald Trump has used social media platforms, such as Twitter, now X, and, more recently, his incongruously named 'Truth Social', to propagate his views. A former White House press secretary and director of communications to President Trump, Stephanie Grisham, confided in her recent speech to the 2024 Democratic National Convention: *"He used to tell me, 'It doesn't matter what you say, Stephanie. Say it enough, and people will believe you'"*[21]. Mr Trump's technique for promoting the false narrative that he won the 2020 Presidential election harks back not only to Plato, but to the chilling reality of the masters of propaganda from last century, Goebbels and Adolf Hitler.

**The impact of artificial intelligence**

In today's world, there are 'alternative facts', 'fake news' and 'misspeaks'. These describe versions of events that do not appear to correspond with reality. Added to these phenomena, is the latest product of the technological age, artificial intelligence or AI. AI is a generic and imprecise description of a computer program that 'learns' or harvests information from large amounts of data and then produces outputs that can appear to emanate from humans.

In the updated guide *AI Decision-Making and the Courts*[22] published by the Australasian Institute of Judicial Administration and the University of New South Wales Faculty of Law and Justice (the AIJA AI Guide), the authors describe 'AI' as *"a broad umbrella term with no single meaning"*[23]. They review various AI systems but describe machine learning as the most well-known sub-field of AI research. Machine learning occurs in a model with parameters that an algorithmic process has set to reflect data or specific experience so that the model 'learns' in order to improve its performance progressively as it processes data or experience. However, as the authors state, such machine 'learning' is not the same as human learning[24]. The authors describe generative AI as an AI system that is capable of generating content in response to prompts, such as the well-known program ChatGPT.

ChatGPT-4 is an example of a large language model or LLM. It produces text in answer to questions or tasks set by a user, such as writing an essay or providing legal research[25]. Importantly, the authors of the AIJA AI Guide say[26]:

> The outputs of generative AI systems might sometimes be true statements, but there is no guarantee that this will be the case based on how these systems function. In particular, there may be no 'truth filter' or source-checking, despite outputs that might suggest otherwise (e.g. "Yes, that is correct"). The term 'hallucinations' is sometimes used to describe outputs that suggest something is the case when it is not or where a non-existent source is cited. Those attributing sentience to tools such as ChatGPT fundamentally misunderstand its nature. There are also significant risks in relying on LLMs.

---

[21] See Samuel Clench, 'Democratic National Convention live: Barack and Michelle Obama to headline day two', *news.com.au* (online, 21 August 2024) < https://www.news.com.au/world/north-america/us-politics/democratic-national-convention-live-barack-and-michelle-obama-to-headline-day-two/live-coverage/2e35d1777d83d627f1439abe194bc916#/entry/21775241>.

[22] Lyria Bennett Moses, Michael Legg, Jake Silove, Monika Zalnieriute, with research assistance from Shahzeb Mahmood, (a joint research project between the Australasian Institute for Judicial Administration and the University of New South Wales Faculty of Law and Justice, December 2023) (AIJA AI Guide).

[23] Ibid at 8 [2.1].

[24] Ibid at 11-14 [2.7].

[25] Ibid at 15-16 [2.8].

[26] Ibid at 16 [2.8] (citations omitted).

AI programs have both uses and flaws that, like the curate's egg, can be good or bad in parts. Worse still, the outputs can have the quality of verisimilitude – they can lie like truth, or in AI jargon, 'hallucinate'. AI can superimpose a person's face on another person's torso to convey a false image online or in a documentary form. It can write essays, letters, webpages, and social media posts that incorporate truthful or accurate material or, sometimes spontaneously, wholly fictitious information that wears the badge of a reliable source.

Much depends, of course, on the algorithms with which the particular AI program is constructed. A chatbot is an AI program that interacts with human users by generating responses that address questions asked or tasks set by the user. So, when Google released its new chatbot, Gemini, last year, users interacting with it began to notice some curious responses. Google had programmed Gemini with instructions that reflected particular ideological values. One such value was designed to promote 'diversity' by avoiding the generation of stereotypes, that other AI tools had used, such as only producing depictions of white men when asked for images of doctors or entrepreneurs. Gemini obeyed its programmed instructions and produced images of a black George Washington and an Asian woman as the Pope. According to an article in *The Economist*[27], Gemini also provided arguments in favour of affirmative action in higher education but refused the user's request to give arguments to the contrary. It also would not write a job advertisement for a fossil fuel lobby group because, *The Economist* related, it told the user *"fossil fuels are bad and lobby groups prioritise the interests of corporations over public well-being"*.

All of the providers of social media and internet search engines use AI algorithms to filter search results, information and what the program characterises as 'news' to match the perceived interest or beliefs of the human user. This can be an effective propaganda tool because it pushes a perspective that matches the recipient's and helps to reinforce what that individual is interested in receiving. Thus, despite the reality that Donald Trump lost the 2020 Presidential election, from the night of the election, he has peddled the lie to his supporters that he had won, and that his victory had been stolen from him. The algorithms filter their internet and social media sources of news with material that reinforces that version of events, and omits contrary information. Mr Trump could not identify any credible basis to demonstrate how the 'theft' occurred but has kept saying that it did ever since. The claim is a lie, not least because about 80 court proceedings brought by him and his supporters seeking to challenge voting outcomes were summarily dismissed for lack of substance. Even his loyal Vice President, Mike Pence, immediate past Attorney-General, William Barr, and most Republican members of Congress, recognised that truth mattered more than propaganda and allowed the election result to stand.

However, as the Chief Justice of New South Wales, Andrew Bell, recently wrote, according to numerous reports, up to 30 million Americans accept Mr Trump's claim. The Chief Justice observed that this statistic impliedly conveyed that a substantial proportion of the American people either was ignorant of the court rulings or simply did not accept them[28]. Those people interact as a group with propaganda peddled by Mr Trump and his supporters in a social media echo chamber, consciously or unconsciously, unaware and unwilling to examine the objective reality. They are in a world where that which does not conform to their perception is labelled 'fake news'.

As I have explained above, propaganda has been used since at least the time of Plato to shape society's, or a section of society's, perception that reality is a falsehood, and a falsehood, reality. Social media and the internet have made it far easier to communicate both truth and misinformation immediately to a vast audience.

---

[27] "Is Google's Gemini chatbot woke by accident, or by design?" (28 February 2024).

[28] Andrew S. Bell, 'Truth Decay and its Implications for the Judiciary: An Australian Perspective' (Speech, 4th Judicial Roundtable, Durham University, 23-26 April 2024) at [14].

**AI in the courtroom**

AI tools are being developed with their creators' ever-increasing claims of accuracy, reliability and efficiency. The authors of a recent study *Artificial Intelligence as Evidence*[29], one of whom was a United States District Court Judge, found that it is very difficult to build AI tools that will operate in a verifiably fair, unbiased manner. This is partly because the concept of fairness, like beauty, is in the eye of the beholder. Society would need to agree a common standard to which a tool must perform, where the setting of that standard involves many value based alternative measures. Human mental processes are affected by unconscious biases and instinctual responses, as Daniel Kahneman showed in *Thinking, Fast and Slow*[30]. He explained one form of bias is conveyed implicitly by the way we frame a question to suggest an answer.

Of course, AI tools can be useful and result in saving time and cost, provided that the human user reviews and checks the product. Chief Justice Bell wrote[31]:

> In September 2023, Lord Justice Birss reported to a Law Society conference that he had tried to use ChatGPT to provide a summary of an area of law and referred to it as "jolly useful" and as having "real potential".

His [Lordship] went on to say:

> "I'm taking full personal responsibility for what I put in my judgment, I am not trying to give the responsibility to somebody else. All it did was a task which I was about to do and which I knew the answer and could recognise as being acceptable."

Persons engaged in litigation have used AI to produce evidence, legal submissions that include fallacious authorities and also character references. Some of these uses raise real concerns about the future integrity of the judicial process. AI has become pervasive in other areas too, leading some universities to require students to return to handwriting exam papers as was the only option for my generation and earlier ones.

In the United States, AI tools have been used in State Courts in relation to quantifying the risk of reoffending for the purpose of criminal sentencing. Giving the reasons of the plurality of the Supreme Court of Wisconsin in *State v Loomis*[32], Bradley J explained the background to this, saying that the Conference of Chief Justices and the American Bar Association had expressed support for efforts by States to adopt policies and risk assessment tools that were effective in reducing recidivism. In *Loomis*[33] the sentencing judge took into account, in order *"to corroborate"* his finding[34], three scores for pretrial, general and violent recidivism, that a proprietary AI program called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) generated. Those scores classified Mr Loomis as presenting a high risk of recidivism.

The Court found that the owner of COMPAS had never disclosed its trade secret as to how the program determined the risk scores or weighed the input factors to arrive at those scores[35]. Disturbingly, Bradley J recognised that *"we are not in a position to evaluate or opine*

---

[29] Judge Paul W. Grimm, Maura R. Grossman and Gordon V. Cormack, 'Artificial Intelligence as Evidence' (2021) 19(1) *Northwestern Journal of Technology and Intellectual Property* 9 at 45-47.
[30] (2011, Penguin).
[31] Andrew S. Bell, 'Truth Decay and its Implications for the Judiciary: An Australian Perspective' (Speech, 4th Judicial Roundtable, Durham University, 23-26 April 2024) at [141] (citations omitted).
[32] 881 NW 29 749 at 752.
[33] Ibid.
[34] Ibid at 757 [28].
[35] Ibid at 761 [51].

*on the scientific reliability"* of conflicting results of certain validation studies that some States had undertaken to evaluate COMPAS' risk scoring[36]. The Court also found that COMPAS:

(1) produced risk scores based on data for groups of high-risk offenders rather than for the individual being sentenced;
(2) used comparisons with a national sample that had not been cross-validated to the local (Wisconsin) population;
(3) had been questioned in some studies as disproportionally classifying offenders in minority groups as having a higher rate of recidivism; and
(4) had been developed not for use in sentencing but for use by correctional authorities in deciding treatment, supervision and parole issues[37].

The Court held that in arriving at a sentence, a judge could take the COMPAS risk scores of an offender for recidivism into account, but also, somehow, could not use those scores to determine the severity of a sentence or whether the offender could be supervised safely and effectively on parole[38]. The mental gymnastic abilities of Wisconsin judges must be considerable now that AI apparently can help them, with those arcane confines, in the very challenging task of sentencing.

It is troubling that, seemingly, in the United States an offender, indeed anyone, cannot test the integrity or validity of COMPAS or other proprietary AI tools' programming and algorithms that produce risk scores which a sentencing court can use, provided it does not do so 'determinatively', whatever that means. Clearly enough, the decision in *Loomis* required that the risk scores were not to be regarded as 'true', but rather as some information of unverifiable integrity, that resulted from a process that was opaque.

The AIJA AI Guide authors observed that research showed COMPAS' predictive accuracy as *"mixed"* and that its possible bias raised *"very serious issues"*[39]. They pointed out that COMPAS' bias reflected the *human* bias inherent in the data from which the program was trained[40]. In *Wong v The Queen*[41], Gaudron, Gummow and Hayne JJ said of the use by a judge of predictive or statistical data in programs in sentencing an offender:

The production of bare statistics about sentences that have been passed tells the judge who is about to pass sentence on an offender very little that is useful if the sentencing judge is not also told *why* those sentences were fixed as they were.

The AIJA AI Guide authors correctly identified that the variables used in COMPAS did not reflect the important human right of each individual to be equal before the law because of who he or she is and not because of what data sets the person may be part of[42]. They gave a telling example of the impact of COMPAS in a 2013 sentence imposed in Wisconsin on Paul Zilly by Judge Babler. Mr Zilly had been found guilty of stealing a lawnmower and tools that he intended to sell for parts. The prosecution and he agreed to a plea deal recommending he be sentenced to one year's imprisonment and subsequent supervision. At first, Judge Babler gave an indicative sentence of 18 months. But, after the judge viewed the COMPAS risk scores, he said that Mr Zilly's risk of reoffending was *"about as bad as it could be"*, and then rejected the plea deal. Judge Babler sentenced Mr Zilly to two years imprisonment[43]. This suggests that the COMPAS risk scores played a decisive role in that sentence.

---

[36] Ibid at 762 [58] note 29.
[37] Ibid at 768-770 [100].
[38] Ibid at 768 [93]-[94].
[39] AIJA AI Guide at 29-30 [3.4].
[40] Ibid at 30 [3.4].
[41] (2001) 207 CLR 589 at 606 [59] (emphasis in original).
[42] AIJA AI Guide at 50 [4.3.1].
[43] AIJA AI Guide at 28 [3.4].

**Lawyers and AI**

In two recent North American cases, lawyers used ChatGPT to write their submissions. This did not prove to be a good short cut because ChatGPT produced hallucinated material. In *Mata. v Avianca, Inc.*[44] Judge Castel of the United States District Court for the Southern District of New York, was dealing with the defendant airline's argument that Mr Mata's claim was time barred – filed too late – because of an article in the *Montreal Convention* that governs liability for damages on international flights. Mr Mata's lawyers' ChatGPT generated submission purported to support propositions with references to authorities, some with law report citations. The defendant airline's lawyers informed the judge that they could not find the supposedly reported cases and that most of the other cases cited did not support the propositions for which they were called in aid. The judge required Mr Mata's lawyers to file an affidavit annexing the cases.

Things then degenerated further. The lawyers also could not find the cases but then asked ChatGPT to produce them, which, obligingly, it did. However, Mr Mata's lawyers did not tell the court initially that ChatGPT had written their submissions or, literally, produced the authorities. One was supposedly a decision of the Eleventh Circuit Court of Appeals, the legal analysis in which Judge Castel described as *"gibberish"*[45]. His Honor said that the 'judgment' ended abruptly without a conclusion. It had a docket number that was for another matter, and although the Federal Reporter citation of 925 F 3d 1291 created by ChatGPT was a real citation, it was for a different decision by the District of Columbia Court of Appeals[46]. In other words, ChatGPT fabricated the decision, including its docket number and reported citation.

The judge noted that some other decisions quoted in the ChatGPT submission actually existed but did not contain the language quoted or support the propositions for which they were cited. Needless to say, Judge Castel criticised the lawyers' conduct in blindly using the AI tool and pointed out the risk it posed to the administration of justice. The judge imposed sanctions and penalties on Mr Mata's lawyers. However, ChatGPT seems to have escaped any sanctions.

In the Supreme Court of British Columbia divorce case of *Zhang v Chen*[47], the husband's lawyer referred in a submission to two cases suggested by ChatGPT. When the wife's lawyers said that they could not find the cases and asked for copies, the husband's lawyer wrote apologising *"for the incorrectness"* of the cases and added references to four new (and real) cases. When the matter returned to court, the husband's lawyer belatedly confessed that ChatGPT had suggested the two earlier cases, and she had not verified them before providing the submission. Masuhara J found those circumstances *"alarming"* but that they were the result of a mistake rather than a deliberate attempt to mislead the court[48]. His Honour observed that citing a fake case was an abuse of process and tantamount to making a false statement to a court. Both Judge Castel and Masuhara J noted the potential for similar situations to result in a miscarriage of justice.

Perhaps the defaulting lawyers in these two cases would have been more circumspect about using the raw output of ChatGPT if they had done what Justice Perry recently did and asked it factual questions about herself. As she told the Commonwealth Law Conference in February 2023, based on the concocted answers she received, the outputs of such tools *"should be approached with a high degree of caution, if not scepticism"*[49].

---

[44] 22-cv-1461 (PKC) (S.D.N.Y.) (June 22, 2023).
[45] Ibid at [27].
[46] Ibid at [27]-[28].
[47] (2024) BCSC 285.
[48] Ibid at [31].
[49] Melissa Perry, 'The Future of Administrative Decisions' (Speech, Commonwealth Law Conference, February 2023) cited in AIJA AI Guide at 38-39 [3.8] footnote 186.

In a case closer to home, earlier this year, in the Supreme Court of the Australian Capital Territory, Mossop J was confronted, when sentencing an offender, with a character reference containing the curious statement that the nominal author had known his brother *"personally and professionally for an extended period"*[50]. Other peculiarities in the letter's phrasing and content led his Honour to infer that it had been generated by a large language module program such as ChatGPT[51]. Fortunately for the offender, his other references were both genuine and helpful to his Honour.

**The future of AI in litigation**

The above recent cases involving misuse of AI tools raise deeper concerns about the future integrity of evidence. So much original evidence nowadays is already in digital form. It comprises the electronic records and associated metadata that are created on computers, email servers, mobile phones and social media accounts. It is obvious that these can be manipulated, more and more skilfully with the assistance of AI tools, to create a new, but false, historical narrative of one or more records or communications. If the other side in the litigation has a different version, the court may have two apparently plausible but divergent records from which to garner the truth. And, of course, there is the potential of hacking into an opponent's electronic records to manipulate them. Not everyone will have access to an expert who could expose as fraudulent a corrupted electronic record. Yet then it will be for a judge or jury to weigh that evidence, as an 'alternative fact', unaware of whether it is authentic or a lie.

There are also those who see AI tools, such as large learning modules, being used to decide litigious disputes. The problems with making an impartial, transparent and reliably accurate computer program for such a role, that I have discussed above, also threaten the ideal of an impartial trial. Article 10 of the *Universal Declaration of Human Rights* provides[52]:

> Everyone is entitled in full equality to a fair and public hearing by an independent and impartial tribunal, in the determination of his rights and obligations and of any criminal charge against him.

This reflects a deeply embedded value in all societies as to the role of an independent judiciary. If a computer program is empowered to adjudicate a dispute, how is its independence, impartiality and reliability to be ascertained or tested, especially if courts, with human judges, such as those in the United States, refuse to allow examination of programs that make such decisions because they are proprietary? The principle of open justice ought to ensure that no such shield be permitted to prevent the public from being able to examine how such a program is constructed and operates.

Moreover, while consistency of outcome for persons with similar disputes or circumstances is an important value in the principle of the rule of law, each case that comes before a court has its own particular features, including the identity and characteristics of the individuals whose conduct has generated the dispute, civil or criminal. Portia articulated an essential part of our concept of justice in her speech to the Duke in Shakespeare's *Merchant of Venice*[53]:

> The quality of mercy is not strained;
> It droppeth as the gentle rain from heaven
> Upon the place beneath. It is twice blest;

---

[50] *Director of Public Prosecutions (ACT) v Khan* [2024] ACTSC 19 at [40].
[51] Ibid at [39].
[52] see too Article 14 of the *International Covenant on Civil and Political Rights*, and c.39 of Magna Carta: "*No free man is to be arrested, or imprisoned, or disseised, or outlawed, or exiled, or in any other way ruined, nor will we go against him or send against him, except by the lawful judgment of his peers or by the law of the land*".
[53] Act IV, scene 1, line 189-210.

It blesseth him that gives and him that takes:
'T is mightiest in the mightiest; it becomes
The throned monarch better than his crown:
His sceptre shows the force of temporal power,
The attribute to awe and majesty,
Wherein doth sit the dread and fear of kings;


But mercy is above this sceptred sway;
It is enthronèd in the hearts of kings,
It is an attribute to God himself;
And earthly power doth then show likest God's
When mercy seasons justice.

**The impact of social media campaigns on individuals**

Moreover, because electronic communication by texts, emails, the internet or social media platforms is instantaneous, many in society have become accustomed to responding in matching time. Facebook and similar platforms give users a choice of liking or disliking or forwarding a post or item of news or information. Frequently users look at a news item, read it cursorily and then respond almost unthinkingly, without taking the time to ponder or process why or how the particular event or subject matter had occurred. Instead, as a normal human response, the recipient forms and communicates a snap judgment. Were the communication to occur face to face with participants discussing it, one may point out that another's first impression was wrong or that there is a different way of looking at the subject matter.

I noted in a speech I gave in 2017[54] how a badly put or careless tweet could ruin a career. I used the example of a public relations executive, Justine Sacco. She posted a tweet to her 170 Twitter followers before she left London, after a flight from New York, to holiday in South Africa: *"Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!"*

By the time her plane touched down in Cape Town, her tweet had gone viral and been retweeted tens of thousands of times. A photographer snapped her in the arrivals hall. The next day, she was sacked from her employment. Her tweet created international outrage. She was denounced by the anonymous purveyors of social media correctness. Yet, when you think about what she did, it was to write a clumsy, silly joke. Had she said it in conversation with some friends, one or more may have said something discouraging, or she may have experienced an awkward silence or look. That would have made her realise that her comment was, or could be seen as, inappropriate. Or, her listeners may just have had a sense of humour, rather than outrage, and laughed. However, the use of social media generated a huge pile-on of unfavourable reaction to a relatively private communication with devastating consequences for Ms Sacco.

[54] Steven Rares, 'Social Media – Challenges for Lawyers and the Courts' (Speech, Australian Young Lawyers' Conference, 20 October 2017); (2018) 45 Aust Bar Rev 105.

Despite the apparent adoption of non-discriminatory values in its programming of its AI chatbot, Google did not practise that preached lesson on its YouTube platform in the circumstances that confronted me in 2022 when hearing a defamation action brought by the former Deputy Premier of New South Wales, the Hon. John Barilaro MP[55]. This case illustrated the enormous reach of social media postings and the power of the operator of the medium, in this instance Google, to permit their misuse. Of course, Google, which has no legal presence in Australia and insisted on being served with the originating application in California, approached the evaluation of what it allowed to be posted and maintained on YouTube both from the perspectives of how the First Amendment operates on such publications and the revenue it earned from them.

The proceeding involved Mr Barilaro's claim that two videos produced by Jordan Shanks on his friendlyjordies YouTube account defamed him and that other videos Mr Shanks posted had aggravated the damage done to Mr Barilaro's reputation. Google controls the operation and administration of YouTube, and, so, was a publisher of Mr Shanks' posted content. Moreover, Google encouraged and facilitated mass publication of the matters complained of, as well as each of Mr Shanks' subsequent videos and the public comments on those videos[56]. By the time of the trial, Google had abandoned each of its defences that I later found were baseless and aggravated the damage to Mr Barilaro from the publications.

Google claimed that it had policies that it applied to protect persons against racist abuse being posted or allowed to remain on its sites. Based on the evidence I found that[57]:

> … the various videos… are replete with racist, hate filled rants that were calculated to bully and publicly hound Mr Barilaro. Mr Shanks repetitively calls him demeaning names, and posts depictions of him, such as "meatball", "greasy", "greasy little scrotum", "wog", "corrupt" and "Italian" in association with direct references or clear allusions to the mafia. And, he advertised the sale of tee shirts and the bruz key chain [which depicted Mr Barilaro's face above a scrotum] on YouTube as part of his videos.

Of course, when used for the first time or in a limited way, such names or depictions can be humorous. But that was not the effect that Mr Shanks' use of these devices had or was intended to have.

Google's internal YouTube policies had definitions of harassment and cyberbullying as content that features prolonged name calling or malicious insults such as racial slurs based on an individual's intrinsic characteristics. The policies also prohibited hate speech and appeared to have been drafted to reflect the norms in the *International Convention on the Elimination of All Forms of Racial Discrimination*[58] and Pt IIA of the *Racial Discrimination Act 1975* (Cth). I found that[59]:

> Despite the above policies being publicly available, the videos in evidence that I have described… show that Google did not appear to take the application of its policies seriously, no doubt because Mr Shanks was very popular and YouTube publications, such as his, earned Google revenue. Moreover, its policies, that were only internally available, that it sought to keep from public scrutiny in this proceeding, reinforced the same themes, as summarised above. The internal hate speech policy identified that a video that "is dedicated to repeating slurs over and over … is likely to cross the comedic line we set".

There have been many similar instances of social media platforms being used to vilify and bully individuals, including school children with, usually, distorted or false assertions about them that they cannot answer directly or at all in the same medium of publication. Resort to

---

[55] *Barilaro v Google LLC* [2022] FCA 650.

[56] Ibid at [321]-[322], citing *Fairfax Media Publications Pty Ltd v Voller* (2021) 392 ALR 540 at 548 [32] per Kiefel CJ, Keane and Gleeson JJ, 563–564 [104]–[105] per Gageler and Gordon JJ.

[57] Ibid at [324]-[325] (internal references omitted).

[58] opened for signature 21 December 1965.

[59] *Barilaro v Google LLC* [2022] FCA 650 at [332].

court proceedings is not for the faint hearted given the enormous legal costs, technicalities of defamation law and the deep pockets of the 'Big Tech' publishers.

**Conclusion**

The influence of the First Amendment of the Constitution of the United States on the owners of almost all social media platforms outside China and Russia is evident because of their unwillingness or inability to prevent or control their users conveying lies or engaging in bullying or other inappropriate behaviours. This is unlike the situation of the mass media of last century which were legally responsible for, and could decide, what they published. Most traditional media felt constrained by objective reality to stick to the facts, that is, the actual facts. How that has changed today. Of course, media publishers have always been able to convey their opinions and their stances on the stories that make the news. Freedom of the press and of opinion guarantee their right, and that of every other member of the community, to do so.

Our current age has seen the new forms of news media – the internet and social media – evolve in ways that allow (as in the past) the beneficial dissemination of information and opinion and the malign use of propaganda. AI has now created the daunting prospect of machines rather than humans making things up and bringing us closer to confronting a situation like in Stanley Kubrick's classic film, *2001, A Space Odessey* when the computer, HAL, went rogue. We need to be careful for our future as a society. It is essential that humans remain the master of what AI is allowed to do, before our use of, and dependence on, the internet, social media and AI get (completely) out of our control.