

JUSTICE STEPHEN ESTCOURT
ADDRESS TO THE STUDENTS' LEGAL SERVICE 2021
BEWARE THE MARCH OF AI

After flattering me into accepting an invitation to speak to you tonight, Nicholas Rae suggested to my associate that I prepare a speech of about 10 to 15 minutes duration.

Well, I recently had an MRI on my knee which took 15 minutes and I know how intolerably long that can be when you have to be still and silent. So we'll make it 10 minutes and I will backdate the sentence to a minute ago when I was introduced. Someone put the timer on me please.

Social justice is like the rule of law. Often spoken about but not always understood in its wider import.

From your perspective you doubtless, and quite rightly, perceive your role in promoting social justice as involving facilitating access to the *justice system*. However, it also involves exercising your skills as lawyers to ensure that the *justice system* remains fair.

How so, you say? How could that be? Laws are made by a democratically elected government. We live under the rule of law. How can our justice system itself not be fair?

Well, some say, "Beware the Ides of March"... I say,... "Beware the march of AI."

I suspect that sections of the legal profession and the judiciary may still apprehend that Artificial Intelligence in the law is confined to software such as that which assists in identifying and collating relevant documents for discovery in major litigation and such like. That assumption would be wrong at any level. Things have moved apace.

AI has populated the legal landscape in far more diverse and controversial ways than predicting family property adjustments or helping clients draft a will. The quest for autonomous thinking machines for use in judicial decision-making is already on the march.

When Chief Justice Roberts of the United States Supreme Court was asked in 2017,¹ whether he could *foresee* a day when smart machines, driven with artificial intelligence, would assist with courtroom fact finding or, more controversially even, judicial decision making, his Honour responded: "*It's a day that's here* and it's putting a significant strain on how the judiciary goes about doing things."

Indeed, key figures in the AI quest, such as Elon Musk have warned that we should be very careful about artificial intelligence because it poses our biggest existential threat. The late

¹ For a video recorded conversation with the President of the Rensselaer Polytechnic Institute on April 11, 2017.

Stephen Hawking observed that the development of full artificial intelligence could spell the end of the human race.

Others, such as Facebook creator Mark Zuckerberg, do not accept that we will become to robots what dogs are to humans. Zuckerberg believes that AI can be built so that it works for and assists humans without taking control.

Australian thinker Toby Walsh, Professor of Artificial Intelligence at the University of New South Wales, who has been hailed as one of the "rock stars" of Australia's digital revolution, suggests in his excellent book *It's Alive*² that as a society we need to begin making choices as to what we entrust to machines.

Walsh writes that the ultimate message of his book is that AI can lead to good or bad, and that while there are many decisions which we could hand over to machines, only some of them should be, even when the machines can make them better than we do.

Writing in the New York Times in December 2017, Corbett-Davies³, Goel⁴, and González-Bailón⁵, pointed out that in courtrooms across the United States, judges now turn to computer algorithms when deciding whether defendants awaiting trial must post bail or can be released without payment, and that algorithms have also proved useful in informing sentencing decisions when used to identify probationers and parolees at low risk of future violence.

The authors note that the increasing use of such algorithms has prompted warnings about the dangers of artificial intelligence, but argue that research shows that algorithms are powerful tools for combatting the *capricious and biased* nature of human decisions. But are they?

Algorithms are designed by humans, and the fear is that algorithms simply reflect the biases of those who develop them, as well as the biases buried deep in the data on which they are built.

Alarming, AI expert Margaret Mitchell, recently fired by Google as head of its ethical AI unit, when employed at Microsoft Research in 2016, pointed out that only about 10 per cent of AI researchers were women. So bias is a real social justice concern.

Davies et al note, for example, that when the 2017 Pulitzer Prize winning investigative news service, ProPublica, examined COMPAS⁶ computer-generated risk scores in Broward County, Florida, in 2016, it found that black defendants were substantially more likely than whites to be rated a high risk of committing a violent crime if released, even among defendants who ultimately were not re-arrested after release.

² Latrobe University Press, 2017.

³ PhD student at Stanford University.

⁴ Assistant Professor at Stanford and Executive Director of the Stanford Computational Policy Lab.

⁵ Assistant Professor at the University of Pennsylvania.

⁶ An acronym for Correctional Offender Management Profiling for Alternative Sanctions.

They make the point however that it is not just biased algorithms, but broader societal inequalities that are to blame, and argue that it is misleading and counterproductive to blame the algorithm for uncovering real statistical patterns.⁷

The procedural fairness of COMPAS might have been determined by the United States Supreme Court in 2017, but unfortunately was not. *Loomis v Wisconsin*⁸ was a petition made to the Court to overturn a Wisconsin Supreme Court ruling in *State v Loomis*⁹. The case challenged the State of Wisconsin's use of proprietary, closed-source¹⁰ risk assessment software in the sentencing of Eric Loomis to six years in prison after the judge rejected a plea bargain.

The petition alleged that, using such software in sentencing, violated the defendant's right to due process because it prevented the defendant from challenging the scientific validity and accuracy of the test, in essence, how the secret algorithm operated. Loomis also alleged that COMPAS violated due process rights by taking gender and race into account

Hearing the case would have given the Court "the opportunity to rule on whether it violates due process to sentence someone based on a risk-assessment instrument whose workings are protected as a trade secret." ¹¹ The Supreme Court denied Loomis' application for a writ of certiorari on 26 June 2017, thus declining to hear the case.

So, should we trust judicial decision-making to algorithms? Certainly issues of bias need to be addressed, but so too do rule of law considerations such as the design and transparency of the algorithm behind the decision, and the right of an accused person to examine it¹².

Moreover, there is the overarching societal question as to whether decision-making will be enhanced at all by autonomous reasoning and, if it is, whether the process is one that should be handed over to robots.

There is perhaps an inherent risk that once we trust these decisions to AI systems, we become prone to over-reliance on them based on implicit trust that the systems cannot make a mistake.

The executive director of *AlgorithmWatch*¹³ Matthias Spielkamp posited in 2017¹⁴ that if you were imagining designing a system to predict which criminals will reoffend, one option would be to optimize for "true positives," meaning that you would identify as many people as possible who are at high risk of committing another crime. One problem with that approach

⁷ Race was not a data input to the software but ZIP codes were.

⁸ Petition for certiorari denied on June 26, 2017.

⁹ 881 N.W.2d.

¹⁰ Privately owned and patented.

¹¹ Wexler, Rebecca (2017-06-13). "Opinion | When a Computer Program Keeps You in Jail" New York Times ISSN 0362-4331.

¹² Doubtless a more appropriate vehicle than *Loomis* will present itself in this regard sooner or later, or legislation will.

¹³ A US advocacy group that analyses the risks and opportunities of automated decision making.

¹⁴ Writing in the online MIT Technology Review on 12 June 2017.

however is that it tends to increase the number of "false positives", that is, people who will be unjustly classified as likely reoffenders.

Spielkamp says that the dial can be adjusted to deliver as few false positives as possible, but that tends to create more "false negatives", that is, likely reoffenders who slip through and receive a more lenient outcome than warranted.

Raising the incidence of true positives or lowering the incidence of false positives are both ways to improve a statistical measure known as "positive predictive value", that is, the percentage of all positives that are true.

Spielkamp points out that one thing this tells us is that the broader society comprising legislators, the courts and an informed public, need to decide what the community wants these sort of algorithms to prioritise.

Is the primary interest in taking as few chances as possible that someone will breach bail or reoffend? What trade-offs should be made to ensure justice on the one hand and lower the social costs of imprisonment on the other?

Fairness may mean different things to someone unfairly kept in prison and to a society concerned at the rate of recidivism. I am sure you are seeing the social justice issues here!

In his latest book, *2062 The World That AI Created*, Toby Walsh notes¹⁵, that in handing over decisions to these machines we need to think carefully about what we want fairness to mean in a given setting. As he says, the irony is that our *technological* future will not be about technology, but about our *humanity*.

He concludes that no matter which way the dials are set, any algorithm will have biases, as it is making a prediction based on generalised statistics and not on someone's individual situation.¹⁶ But, he argues, we can still use such systems to guide decisions that are wiser and fairer than the ones humans tend to make on their own.

As we will see however, that view may not be universally shared in Australia.

On the one hand, Dr Nigel Stobbs and Professors Dan Hunter and Mirko Bagaric in their article *Can Sentencing Be Enhanced by the Use of Artificial Intelligence*¹⁷, suggest that although sentencing decisions are influenced by more than 200 considerations, sentencing law and practice is on its face amenable to automated decision-making. They argue that most of the relevant facts are established prior to or following a relatively short plea hearing, and it is generally relatively straightforward to identify the relevant sentencing objectives and aggravating and mitigating considerations.

¹⁵ *2062 The World That AI Created* Latrobe University Press, 2018 at p 167.

¹⁶ Focussing on statistics may however compound the external effects over time because an AI system designed to "learn" as it processes more cases will reinforce the biases.

¹⁷ (2017) 41 Crim LJ 261

They conclude that there are "a number of shortcomings" associated with the intuitive synthesis sentencing process of human judges and after examining whether a computerised sentencing process had the capacity to remedy these shortcomings, without introducing significant additional problems, they recommend that sentencing algorithms should be developed and trialled as an adjunct to existing sentencing practices.

If the trial were to be successful they recommend that consideration be given to the wide-ranging use of computer-assisted sentencing decisions.

Conversely, in his paper *Technology and the Law*¹⁸, Justice Geoffrey Nettle (although in the context of open textured reasoning), eloquently observed:

" Given th[e] degree of reticence about allowing judges to make decisions based on broad conceptions of contemporary social contexts to doing justice, it is not unlikely that society would also be resistant to the idea of policy choices being made by a computer on the basis of *a priori* determinations¹⁹ made by a cohort of *unelected, unanswerable and essentially unknown software engineers and legal specialists working alone and largely unexamined* in the development of a database and complex algorithm intended to function as a modern day computational law Atkinian replacement". (Emphasis added.)

In 2062²⁰, Walsh notes that most experts in AI believe that there is a 50 per cent chance that we will have created machines that can think as well as humans by the year 2062. The year that we *Homo sapiens* begin to be overtaken by our successor, *Homo digitalis*.

Law firms and academics are starting to agitate for law schools to modify or abandon the Priestly Eleven core law subjects and the case method of teaching law, and to create new courses in human creativity²¹ and in coding and legal technology.

Nettle J certainly seems to believe that the issues of scepticism, transparency and the mindset of the algorithm developers themselves will result in the skills of counsel and judges having to change. In his paper²², he opines that submissions and judgments may need to include explicit reference to the programs and the results which they recommend. Possibly, he said, there will be competing programs which dictate different conclusions, and counsel and judges may need to analyse each of them and compare them. His Honour offered this view of the future:

"Just as the adoption of robotics in industry is changing the role of tradesmen into skilled computer technicians and industrial plant managers from skilled personnel managers to skilled computer scientists, so would the role of counsel and judges become increasingly one

¹⁸ Presented to the Bar Association of Queensland Annual Conference on 27 February 2016.

¹⁹ This raises the question as to how algorithms could allow for incremental change in sentences reflecting and responding to developments in community understanding of the impact of particular offences as was acknowledged as legitimate by the High Court in *R v Kilic* [2016] HCA 48 at [21].

²⁰ Ibid fn 30.

²¹ Walsh states in 2062 at p123 that the irony is that our *technological* future will not be about technology, but about our *humanity*.

²² Ibid fn 33.

of a skilled computer scientist, with the capacity to identify the limitations in programs and to fashion submissions and judgments about them."

In 2062 Walsh argues that depriving people of their liberty is one of the most difficult decisions we make as a society, and that even if machine learning programs were trained to avoid bias we would hand over an important part of our humanity if we outsourced such decisions to robots.

Whichever view one takes, it is true, as Chief Justice Roberts observed, that the day is upon us and we need to address the issue of thinking machines in the judicial process now.