Thinking Machines and Smiley Faces

Justice Stephen Estcourt AM and Ms Karen Marr*

The reality of autonomous thinking machines for use in judicial decision-making is already upon us. The potential exists now for artificial intelligence tools to be provided to the judiciary to aid in the setting of parole periods, the granting or refusing of bail and in sentencing decisions. However, algorithms are designed by humans, and they reflect the biases of those who develop them, as well as the biases buried deep in the data on which they are built. Society needs to make decisions now as to what outcomes it wants algorithms to prioritise. Educating the public to be sceptical about algorithm results, ensuring transparency and auditing algorithms for bias, as well as establishing regulatory frameworks by legislation, will all go towards ensuring that some legal outcomes can be as fairly determined by a machine as by a judge. But will that simply give us another competent judge?

In 1950, four years before he took his own life at the age of 41, following chemical castration after being found guilty of homosexual acts for which he was posthumously pardoned, Alan Mathison Turing predicted that by the end of the 20th century one would "be able to speak of machines thinking without expecting to be contradicted". Turing of course was the genius who engineered one of the first ever computing devices, the Bombe, which was most famously used in 1941 to assist the breaking of the Enigma Code.

Two-thirds of a century later the Saudi-backed Soft Bank Vision Fund offered no such contradiction, when in 2018 it committed US$100 billion to investment across a wide range of technology sectors, including but not limited to "the Internet of Things",[1] artificial intelligence (AI) and robotics. Other major corporations such as Facebook, Google, Toyota and OpenAI[2] are spending billions of dollars on developing autonomous thinking machines. IBM alone has spent US$1 billion developing *Watson*, a question-answering computer system capable of answering questions posed in natural language.

We suspect that sections of the legal profession and the judiciary may still apprehend that AI in the law is confined to software such as that which assists in identifying and collating relevant documents for discovery in major litigation. Such an assumption would be wrong at any level. Things have moved apace. In recent times the Family Court of Australia has trialled an AI system called *Split Up* which predicts the distribution of assets after separation and, in November 2017, a Darwin legal practice commissioned a client-friendly system known as *Ailira.*[3]

Ailira can assist clients with consumer legal advice from wills to business structuring and asset protection. With a few clicks of a mouse clients can enter their details and will then be asked a few simple questions by *Ailira*, before the robot generates a completed last will and testament.

But, as will be seen, AI has populated the legal landscape in far more diverse and controversial ways than predicting property adjustments or helping clients draft a will. The quest for autonomous thinking machines for use in judicial decision-making is already well underway.

In 2017, an experiment was ceased by Facebook when it was discovered that two AI programs were communicating with each other in a new language which they had created and exclusively understood. The AI programs successfully negotiated an agreement between themselves using this bizarre language.

---

* Justice Estcourt: Judge of the Supreme Court of Tasmania. Karen Marr: Assistant Manager of the Court responsible for the strategic development of ICT. This article is the text of a speech delivered at the Supreme and Federal Court Judges' Conference in Hobart in January 2019.

[1] An inextricable mixture of hardware, software, data and service, known as IoT – eg Google Home and Amazon Alexa.

[2] A non-profit AI research company, discovering and enacting the path to safe artificial general intelligence.

[3] An acronym for Artificially Intelligent Legal Information Resource Assistant.

However, in early 2018, Google allowed its AI-based Translate tool to continue in its organic development, enabling the translating of things into and out of the newly developed language.[4]

The potential exists now for AI tools to be provided to the judiciary to aid parole, bail and sentencing decisions. Such tools would be developed and supported by the bureaucracy. Therefore, as will be seen accountability of the methodologies used to form the algorithms would need to be considered should a challenge to the tool arise.[5]

When Chief Justice Roberts of the United States Supreme Court joined the President of the Rensselaer Polytechnic Institute in 2017,[6] he was asked whether he could foresee a day "when smart machines, driven with artificial intelligence, will 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 AI as it poses our biggest existential threat. The late Stephen Hawking observed that the development of full AI could spell the end of the human race.

Others, such as Facebook creator Mark Zuckerberg, widely touted as a 2020 United States Presidential candidate with the data and algorithmic capacity to win, 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*[7] 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, Sam Corbett-Davies,[8] Sharad Goel,[9] and Sandra González-Bailón,[10] 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 AI, but argue that research shows that algorithms are powerful tools for combating 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.

The authors note, for example, that when the 2017 Pulitzer Prize–winning investigative news service, ProPublica, examined COMPAS[11] 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. They make the point however that it is not just biased algorithms, but broader societal inequalities

[4] "Facebook's artificial intelligence robots shut down after they start talking to each other in their own language", *The Independent*, 31 July 2017.

[5] Moses Bennet, "Artificial Intelligence in the Courts, Legal Academia and Legal Practice" (2017) 91 ALJ 561.

[6] For a video recorded conversation on 11 April 2017.

[7] Toby Walsh, *It's Alive* (La Trobe University Press, 2017).

[8] PhD student at Stanford University.

[9] Assistant Professor at Stanford and Executive Director of the Stanford Computational Policy Lab.

[10] Assistant Professor at the University of Pennsylvania.

[11] An acronym for Correctional Offender Management Profiling for Alternative Sanctions.

that are to blame, and argue that it is misleading and counterproductive to blame the algorithm for uncovering real statistical patterns.[12]

The procedural fairness of COMPAS might have been determined by the US Supreme Court in 2017, but unfortunately it was not. *Loomis v Wisconsin*[13] was a petition made to the Court to overturn a Wisconsin Supreme Court ruling in *State v Loomis*.[14] The case challenged the State of Wisconsin's use of proprietary, closed-source[15] 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.

As Wikipedia observes, 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".[16] The Supreme Court denied Loomis' application for a writ of certiorari on 26 June 2017, thus declining to hear the case.

Another interesting example of how bias can affect the utility of AI is provided by the demise of Google Flu Trends which used the geographical trends of Google search queries for symptoms and cures to predict where influenza might be breaking out. As Google's search algorithm became smarter it would suggest the search query before the user had finished typing, thus introducing a biased offering and skewing the integrity of the inquiry. It was discontinued in 2013, only five years after it began.

Yet another example can be found in Microsoft and Google translation services whose algorithms are based on learning from sources replete with gender bias. This results in sexist translations derived from historic stereotypes as to the roles and traits and occupations of women and men in society.

So what are algorithms, how do they work, what can be done to eliminate bias within them, and should we trust judicial decision-making to them?

Wikipedia tells us that an algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function. Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state.

More simply, and somewhat ironically, the children's website Tynker explains that an algorithm is a detailed step-by-step instruction set or formula for solving a problem or completing a task. In computing, programmers write algorithms that instruct the computer how to perform a task.

Tynker explains that when you think of an algorithm in the most general way, they are everywhere. A recipe for making food is an algorithm as is a method for folding shirts or even your morning routine.

Computer algorithms surround us in our daily lives. Facebook's news feed algorithm considers 100 or so variables or "data points" about its users, such as photographs, locations visited, conversations had, videos viewed, comments made and so on, and can predict whether a user is likely, for example, to "Like" a particular post. The algorithm predicts that likelihood and that prediction is quantified as a "relevancy score" specific to both that user and the particular post.

Facebook advertisements are also given relevancy scores so that users only see the advertisements that the algorithm predicts as being important to them.

[12] Race was not a data input to the software but ZIP codes were.

[13] *Loomis v Wisconsin*, 137 S Ct 2290 (2017). Petition for certiorari denied on 26 June 2017.

[14] *State v Loomis*, 881 NW 2d 749.

[15] Privately owned and patented.

[16] Rebecca Wexler, "Opinion | When a Computer Program Keeps You in Jail", *New York Times*, 13 June 2017, ISSN 0362-4331.

Over 75% of all internet[17] websites contain hidden Google trackers, and some 25% have hidden Facebook trackers. So Google and Facebook are watching us on many of the websites we visit, as well as when we use their search or social media platforms.[18]

Google and Facebook also use our data as input for AI algorithms that put us in a personal digital silo that dictates what we individually see on their platforms. That explains why if you have been searching on the internet to plan your next holiday you might suddenly see a cruise ship advertisement appear on your Facebook news feed or Google page. Or eerier still, it explains why, if you have actually physically visited a travel agent, the same thing occurs.[19] Even your Google Home or Amazon Alexa monitor, or your Fitbit wristband, if you have one, all record your data[20] and transmit it to the Cloud[21] as food for greedy algorithms.

Until 2017 Google's servers read users' incoming emails and used the information detected by the algorithm to target you for advertising. At least that has now ceased. However, as differently skilled algorithms can be programmed to read millions of emails, monitor millions of phone calls and view millions of CCTV cameras all at the same time, the need for vigilance remains.

The increasing expenses associated with Cloud-based data storage perhaps might prove a natural brake on massive data retention by ensuring that upon an algorithmic outcome being achieved all data is purged.[22]

Returning to the question earlier posed as to whether we should trust judicial decision-making to algorithms, 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 it.[23] Moreover, there is the overarching question as to whether decision-making will be enhanced at all by autonomous reasoning and, if it is, whether the process should be handed over to robots. There is perhaps an inherent risk that once we trust these decisions to AI systems, we thereby become prone to over-reliance on them based on implicit trust that the systems cannot make a mistake.

AI, coupled with blockchain technology,[24] may be an antidote to justifiable public distrust.[25] As information passes between parties it is encrypted and decrypted, but moreover it maintains the integrity of each transaction or decision step performed because of its immutability born of the wide distribution of the ledger recording those steps, which cannot be altered without consensus.[26]

Blockchain would offer information that is unable to be interfered with, while displaying transparency of de-identified decision-making processes between two parties. Blockchain treats each transaction as a block which, upon verification, is placed on the end of a sequential chain of all other transactions. Thus, it could facilitate decision-making process audits.[27]

---

[17] There is an ongoing debate as to whether the noun "internet" should be capitalised. We choose not to.

[18] Safari and Firefox will prompt you to be aware of this if you are using either of them as your web browser, but Google Chrome will not.

[19] If location services are enabled on your mobile telephone.

[20] And sometimes your private conversations.

[21] A metaphor for the internet.

[22] ZD Net, *The Paradox of Cloud Data*, 16 December 2016 <https://www.zdnet.com/article/the-paradox-of-cloud-data-it-saves-money-but-can-be-costly>.

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

[24] A blockchain is a decentralised, distributed and public digital ledger that is used to record transactions across many computers so that the record cannot be altered retroactively without the alteration of all subsequent blocks and the consensus of the network.

[25] Marc Pilkington, "Blockchain Technology: Principles and Practice" in F Xavier Olleros and Majlinda Zhegu (eds), *Handbook of Research on Digital Transformations* (Edward Elgar, 2016).

[26] Robert Size, "Taking Advantage of Advances in Technology to Enhance the Rule of Law" (2017) 91 ALJ 575.

[27] Bernard Miller, "Artificial Intelligence and Blockchain: 3 Major Benefits of Combining These Two Mega-Trends", *Forbes*, 2 March 2018.

In criminal courts where it is often necessary to collect and reimburse bail deposits for example, blockchain smart contracts could be used to access the payees' bank accounts and reimburse them directly upon the sureties' release.

Returning to the issue of bias, after ProPublica's investigation, Northpointe Inc, the company that developed COMPAS, argued that the data had been misinterpreted. Ironically, perhaps predictably, algorithm experts assert that both sides of the argument can be correct, depending on the measure used as to what is fair.

The executive director of *AlgorithmWatch*[28] Matthias Spielkamp posited recently[29] that if you were imagining designing a system to predict which criminals will reoffend, one option would be to optimise 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 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.

ProPublica compared false positive rates and false negative rates for blacks and whites and found them to be skewed in favour of whites. Northpointe Inc, by contrast, compared the positive predictive values for different races and found them to be similar. In part because the recidivism rates for blacks and whites do in fact differ, it is mathematically likely that the positive predictive values for people in each group will be similar, while the rates of false negatives are not.

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. In his latest book, *2062 The World That AI Created*, Toby Walsh notes,[30] in handing over decisions to these machines we need to think carefully about what we want fairness to mean in a given setting.

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.[31] 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,*[32] 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 because 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 appropriate sentencing objectives and aggravating and mitigating considerations.

---

[28] A US advocacy group that analyses the risks and opportunities of automated decision making.

[29] Writing in the online MIT Technology Review on 12 June 2017.

[30] Toby Walsh, *2062 The World that AI Created* (La Trobe University Press, 2018) 167.

[31] Focusing 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.

[32] Dr Nigel Stobbs, Dan Hunter and Mirko Bagaric, "Can Sentencing Be Enhanced by the Use of Artificial Intelligence" (2017) 41 Crim LJ 261.

They conclude that there are "a number of shortcomings" associated with the intuitive synthesis sentencing process. After examining whether a computerised sentencing process had the capacity to remedy those 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*,[33] 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* determinations[34] 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.

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. In *2062*,[35] Walsh notes that most experts in AI believe that there is a 50% 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.*

In their article *An Intelligence in Our Image, The Risks of Bias and Errors in Artificial Intelligence*,[36] Osonde Osaba and William Welser IV argue that combating algorithmic bias would benefit from an educated public capable of understanding that algorithms can lead to inequitable outcomes,[37] but that in any event, if we are to rely on algorithms for autonomous decision-making, they need to be equipped with tools for auditing the causal factors behind key decisions. This of course begs the question of whether the owners of the particular AI systems fully understand how or why they function.

Algorithms that can be audited for causal factors, the authors claim, can give clearer accounts or justifications for their outcomes and this, it is said, is especially important for understanding statistically disproportionate outcomes.

The authors also note that while scepticism and transparency are desirable measures, the technical research on bias in machine learning[38] and AI[39] algorithms is still in its infancy.[40] They observe that questions of bias and systemic errors in algorithms also demand a different kind of wisdom from algorithm designers and data scientists. Those practitioners, it is said, are often engineers and scientists with less exposure to social or public policy questions, and their demographics are often less than diverse. As

[33] Justice Geoffrey Nettle, "Technology and the Law" (Presented to the Bar Association of Queensland Annual Conference on 27 February 2016).

[34] 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) 259 CLR 256, [21].

[35] Walsh, n 30.

[36] Osonde Osaba and William Welser IV, *An Intelligence in Our Image, The Risks of Bias and Errors in Artificial Intelligence* (Rand Corporation, 2017).

[37] The authors cite the apparently routine questioning by dating site users of the results of date matching algorithms as an example of growing informed cultural scepticism.

[38] "Machine learning" refers to AI that is based around the idea that we should be able to give machines access to data and let them learn for themselves. ("Deep learning" is a subset of machine learning that has networks capable of learning unsupervised from data that is unstructured or unlabelled. Also known as Deep Neural Learning or Deep Neural Network it seeks to mimic an infant's brain).

[39] Artificial Intelligence is the broader concept of machines being able to carry out tasks in a way we would consider "smart".

[40] Some machine learning systems are difficult to shift from their initial programming without building the system over again. They can be little more than "black boxes" which are fed data from which they learn and the results are analysed to see if they fit the expected outcome.

they make myriad design choices, some of which may have far-reaching consequences, diversity in the ranks of algorithm designers could help to improve sensitivity to potential disparate impact problems. Alarmingly, Margaret Mitchell when an AI expert at Microsoft Research in 2016, pointed out that only about 10% of AI researchers are women.

In a similar vein, 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 creativity[41] 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 paper, already referred to,[42] he opined 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 of a skilled computer scientist with the capacity to identify the limitations in programs and to fashion submissions and judgments about them.

Osaba and Welser also conclude that the drive to remedy algorithmic bias should be tempered with a healthy dose of regulatory restraint. That is undeniable.

Legislation will be needed to regulate the choices society wishes to be made about the algorithmic make-up of robotic judges. There has not to date been any legislation anywhere in the world that protects citizens from robots per se. Asimov's Laws[43] provided that:

(1) A robot may not injure a human being or, through inaction, allow a human being to come to harm.
(2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
(3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws
(0) A robot may not harm humanity, or, by inaction, allow humanity to come to harm.[44]

However, as the European Parliament points out,[45] Asimov's Laws must be regarded as being directed at the designers, producers and operators of robots, including robots assigned with built-in autonomy and self-learning, since those laws cannot be converted into machine code.

In its 2017 resolution with recommendations to the Commission on Civil Law, the European Parliament stated in the preamble:[46]

> A. whereas from Mary Shelley's *Frankenstein's Monster* to the classical myth of Pygmalion, through the story of Prague's Golem to the robot of Karel Čapek, who coined the word, people have fantasised about the possibility of building intelligent machines, more often than not androids with human features;
>
> B. whereas now that humankind stands on the threshold of an era when ever more sophisticated robots, bots, androids and other manifestations of artificial intelligence ("AI") seem to be poised to unleash a new industrial revolution, which is likely to leave no stratum of society untouched, it is vitally important for the legislature to consider its legal and ethical implications and effects, without stifling innovation;

---

[41] Walsh states in *2062* at 123 that the irony is that our *technological* future will not be about technology, but about our *humanity*.

[42] Nettle, n 33.

[43] Isaac Asimov, "Runaround", *Astounding Science Fiction* (Street & Smith, March, 1942).

[44] Asimov claimed in 1942 that these laws came from the *Handbook of Robotic*s 56th ed published in 2058 (sic). The zeroth law (0) was a later addition said to take precedence over the original three laws. But Walsh asks "how can a robot decide what will harm humanity" as the dilemma posed by an autonomous vehicle choosing which human to kill or injure (Springer only published the 2nd edition of their *Handbook of Robotics* in 2008 ☺).

[45] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

[46] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics.

The resolution recommendations include that laws should stress that the development of robot technology should; complement human capabilities and not replace them; consider it essential, in the development of robotics and AI, to guarantee that humans have control over intelligent machines at all times; consider that special attention should be paid to the possible development of an emotional connection between humans and robots – particularly in vulnerable groups (children, the elderly and people with disabilities) – and highlight the issues raised by the serious emotional or physical impact that this emotional attachment could have on humans.

As far as we are aware however,[47] at that level of abstraction at least, Asimov's Laws presently remain the iconic touchstone. Across the board regulation is lacking. Indeed it was only as recently as 2018 that serious attempts to simply protect individuals' data and to require transparency in algorithmic decision-making commenced.

The failure to foresee the need for a prophylactic legislative framework has been much lamented virtually only since 2018 when the Facebook/Cambridge Analytica scandal jolted society into an understanding of data mining and the associated loss of privacy.[48] Widespread, if belated, discussion on data protection legislation followed that incident.

The latest report of the House of Lords Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?*,[49] notes that AI is a tool already deeply embedded in our lives, and warns that the prejudices of the past must not unwittingly be built into automated systems. The report suggests that to do this means not only using established concepts, such as open data[50] and data protection legislation, but also the development of new frameworks and mechanisms, such as data portability[51] and data trusts.[52]

The *Data Protection Act 2018* (UK) replaces the previous Act of the same name that had been in place since 1998 and supplements major reforms to data protection laws that are contained in the 2018 EU *General Data Protection Regulation* (GDPR). The GDPR has direct effect in the United Kingdom and in other European Union member states, and was applied from 25 May 2018. The new *Data Protection Act 2018* contains provisions will which allow for continuation of the GDPR in the United Kingdom, post-Brexit.

The 2018 Act provides that the Information Commissioner's Office (ICO) has statutory duties to produce a number of new codes of practice in areas such as data sharing, direct marketing, and the processing of personal data by journalists, as well as in relation to age-appropriate design of websites, apps and other "information society services" likely to be accessed by children.

Described as a "Copernican Revolution"[53] in data protection law by highlighting the need for algorithms to be non-discriminatory, transparent and open to human interpretation, Art 22 of the GDPR enshrines a right to explanation whereby individuals can require an explanation of an algorithmic decision made about them.[54] The GDPR has provisions on both automated individual decision-making (making a decision solely by automated means without any human involvement) and profiling (automated processing of

[47] Leaving aside para 71 and Art 22 of the *2018 EU General Data Protection Regulation* (GDPR).

[48] By its own admission Cambridge Analytica gathered personal information on Facebook users without their knowledge through surveys. It claimed to be able to pinpoint voters and turn the tide at the ballot box in favour of its customers by using up to 5,000 data points it had on 230 million American voters.

[49] Report of Session 2017–2019.

[50] "Open data" is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright or patents.

[51] The right to data portability allows individuals to obtain and reuse their personal data for their own purposes across different services.

[52] These trusts are not a legal entity or institution, but rather a set of relationships underpinned by a repeatable framework, compliant with parties' obligations to share data in a fair, safe and equitable way.

[53] Goodman and Flaxman, "2016 ICML Workshop on Human Interpretability in Machine Learning" (WHI 2016), New York, NY.

[54] ICO Guide to the General Data Protection Regulation (GDPR). Individual rights.

personal data to evaluate certain things about an individual).[55] Profiling can be part of an automated decision-making process.

The GDPR applies to all automated individual decision-making and profiling. Article 22 contains rules to protect individuals if they are subject to solely automated decision-making that has legal or similarly significant effects on them. This type of decision-making can only be carried out where the decision is:

- necessary for the entry into or performance of a contract; or
- authorised by Union or Member state law applicable to the controller; or
- based on the individual's explicit consent.

Algorithm users must state whether any of their processing falls under Art 22 and, if so, must:

- give individuals information about the processing;
- introduce simple ways for them to request human intervention or challenge a decision;
- carry out regular checks to make sure that their systems are working as intended.

In these important areas Australian legislation barely scrapes the surface. The *Privacy Act 1988* (Cth) is concerned only with "personal information" defined as information or an opinion about an identified individual, or an individual who is reasonably identifiable.

In 2018 data protection measures were introduced by the insertion into the *Privacy Act* of Pt IIIC. However that part is only concerned with the notification of eligible data breaches which involve access, disclosure or loss that causes serious harm to a person.[56]

So, deciding as a society which false positives and which false negatives we want algorithms to prioritise, educating the public to be sceptical about algorithm results and ensuring transparency as well as auditing algorithms for bias and establishing regulatory frameworks by legislation, will all go towards ensuring that some legal outcomes can be fairly determined by a machine as well as a judge. But what will that give us but another competent judge? Perhaps a better judge in some cases but perhaps not. Almost certainly it will be one that counsel and human judges will scrutinise, as Nettle J adumbrated, either at first instance or on appeal, or both.

The polemic we referred to at the beginning of this article posed by Walsh remains to be answered. Human judges can and do explain the reasons for their decisions. To date at least approaches to AI such as deep learning systems cannot. Nor do black boxes have the consciousness to ponder such reasons ethically.[57] But even if it were otherwise, the question remains that while there are many decisions which we could hand over to machines, should we? Or should only some of them be, even if the machines could make them better than we do?

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.

Writing in an online academic publication,[58] Cambridge PhD Candidate[59] Christopher Markou suggests that it would be a terrible idea to hand over judicial decision-making to machines and that we should not do so.

He asks, "what is the point of offloading decision making to an algorithm on matters that are uniquely human? Why do we go through the trouble of selecting juries composed of our peers?"

In his view the standard in law has never been one of perfection, but rather the best that our abilities as humans allow us. "We make mistakes but, over time, and with practice, we accumulate knowledge on how not to make them again – constantly refining the system" he says.

---

[55] Paragraph 71 of the preamble to the GDPR, which is not itself law, requires data controllers to prevent discriminatory effects of algorithms processing sensitive personal data (the latter defined in Art 9).

[56] Section 26WE.

[57] The problem of ethical dilemmas created by decisions required of autonomous vehicles has only recently entered the public discourse. See <http://moralmachine.mit.edu/>.

[58] The Conversation, 16 May 2017.

[59] Faculty of Law of the University of Cambridge.

Markou argues that COMPAS, and systems like it, represent the "black boxing" of the legal system, and that this must be resisted forcefully, as legal systems depend on continuity of information, transparency and ability to review.

One might be tempted to agree, given that researchers from the University of Maryland and Dartmouth College have developed an AI system called Deception Analysis and Reasoning Engine (DARE) that can detect facial micro-expressions such as eyebrows raising, frowning, lips protruded, lip corners turning up, and head side turn, to establish if a person is lying.

DARE was trained by using visual recordings of witnesses cross-examined in courtrooms. It is said to be vastly superior to humans in determining dishonest behaviour.

Systems like DARE could provide judges and jurors with assistance in determining the guilt of an accused, or the veracity of a witness' evidence.

Visual recording technology is already widely adopted within courtrooms around the world, and in some jurisdictions it is specifically legislated for in order to capture high-quality visual recordings in the case of special and child witnesses. The researchers of DARE note that the system could become much more powerful if its recordings were supplemented with audio recordings and transcripts.[60]

What then is the answer to Dr Walsh's riddle? How do we allay Stephen Hawkins' and Elon Musk's fear of existential threat? We, the authors, do not purport to have the answer. Accordingly, we thought that we might leave you with a word or two about emoji to cheer you up. ☺

In their 2017 paper *The Emoji Factor: Humanizing the Emerging Law of Digital Speech*,[61] Deakin Law School researchers Dr Elizabeth Kirley and Associate Professor Marilyn McMahon examined emoji in criminal, tort and contract law contexts and found that they are being progressively recognised, not as a joke or ornament, but as the first step in "a non-verbal digital literacy with potential evidentiary legitimacy to humanize and give contour to interpersonal communications".

The authors write that while it is early days to assess its linguistic and social value, the emoji phenomenon has triggered an emerging academic literature aimed at studying the icons as components of a discrete digital language with evidentiary status. They examine a number of criminal, contract and defamation cases and identify some "interpretative challenges that arise when traditional legal doctrine and procedure are applied to emoji laden content".

In the area of criminal law they note that emoji took a significant step towards legal legitimacy with the high profile trial of Ross Ulbricht in 2015. Ulbricht was the creator of Silk Road, an online illicit drug marketing enterprise investigated in the United States for over $200 million in illegal drug sales.

At the trial the prosecutor read into evidence the text of an internet post created by Ulbricht, without referencing the included smiley-faced emoji. The text read "I'm so excited and anxious for our future, I could burst ☺". The trial judge subsequently instructed the jury to incorporate the emoji in their deliberations of the accused's intentions.

Another criminal case involved a young student from Virginia whose Instagram posts were intercepted by police in 2015 following concerns that the combination of text and a gun, knife, and bomb emoji, and their placement next to each other, conveyed a credible threat of violence. The student was charged with computer harassment and threatening school personnel.

In a contract case from Israel a couple conducting a messaging exchange with a landlord concerning a property he had listed for rent included a message with a string of emoji (a smiley face, a comet, a champagne bottle, dancing figures and more) interspersed with an expression of interest and questions about setting up a viewing time. The landlord subsequently removed the listing, relying on what he believed was a firm contract. The couple then stopped returning the landlord's messages. He sued, claiming that he had relied on the messages to indicate consensus. In a small claims court, the judge

[60] "The robot that knows when you're lying: Scientists create an AI that can detect deception in the courtroom (and it's already 'significantly better' than humans)", *The Daily Mirror Australia*, 20 December 2017.

[61] Elizabeth Kirley and Marilyn McMahon, *The Emoji Factor: Humanizing the Emerging Law of Digital Speech*, Tennessee Law Review, April 2018.

relied on the defendants' repeated expressions of interest, their misleading messages with festive icons, and a smiley face at the end of the negotiations, to find for the plaintiff.

In the area of tort law the authors write that given, the impulsive nature of social media and the possibility of immediate and widespread dissemination, it is unsurprising that emoji have featured in several defamation cases and claims for intentional infliction of emotional distress.

Sending nearly naked selfies and sexually explicit messages clearly raise the potential for various claims they write, but a recipient who responds to semi-naked photographs by informing the sender that he or she is missed and embellishing the text message with an emoji blowing a kiss, is likely to find an unsympathetic court.

The first Australian case to feature full colour emoji (a heart and an angry face) involved the grant of injunctive relief pending the trial of allegations of breach of contractual undertakings, defamation and harassment in September 2018 in *School for Excellence Pty Ltd v Trendy Rhino Pty Ltd*.[62] It is unlikely to be the last.

[62] *School for Excellence Pty Ltd v Trendy Rhino Pty Ltd* [2018] VSC 514, [25].