JUDICIAL DECISION-MAKING AND EXPLAINABLE ARTIFICIAL INTELLIGENCE

A Reckoning from First Principles

In light of rapid developments in legal technology, it is timely to begin considering whether, and if so how, artificial intelligence ("AI") can replace judges. However, given that law plays a crucial role in maintaining societal order, that judges are a crucial part of ensuring the continued well-functioning of the law, and also that there are still many unknowns in the use and deployment of AI, it would be prudent to examine and understand exactly what roles judges play in the legal system, and how they do so, before we make any bold steps towards replacing judges with AI. This article examines the current and reasonably foreseeable state of AI to consider its capabilities, as well as the process by which judges make decisions and the duties they are subject to. This article will then consider whether or how AI, given its current and foreseeable state of development, may be used in judicial decision-making, and what safeguards may be required to ensure continued confidence in a well-functioning justice system.

Shaun **LIM**[1]
LLB (Hons) (National University of Singapore);
Advocate and Solicitor (Singapore);
Research Assistant, Centre for Technology, Robotics, Artificial Intelligence & the Law, Faculty of Law, National University of Singapore

I. Introduction

1 Advancements in artificial intelligence ("AI") techniques with demonstrable results have led to a boom in AI research, development and marketing, especially in fields dominated by specialist professionals whose knowledge was thought to be impossible for AI to replicate, such as in medicine and law.[2] In law itself, AI solutions are being

1 This article expands upon a prior research paper for credit in the course of the author's LLB degree. The author is indebted to Assoc Prof Daniel Seng for his comments and suggestions on this article, as well as to District Judge Wong Li Tein for her comments on the earlier version of this article.

2 See, *eg*, Xiaomin Mou, "Artificial Intelligence: Investment Trends and Selected Industry Uses" (EMCompass Note 71, September 2019) <https://www.ifc.org/ *(cont'd on the next page)*

commercially offered for applications such as due diligence, legal analytics, outcome predictors, document automation and practice management.[3] The Singapore judiciary, long an enthusiastic supporter of technological solutions in the courts, has itself pioneered many legal technological innovations over the years, such as the Integrated Case Management System[4] for criminal case management, the Automated Court Documents Assembly[5] service for certain categories of cases such as bankruptcy and Magistrate's Complaints, the Community Justice and Tribunals System[6] for the Small Claims Tribunal, Community Disputes Resolution Tribunal and Employment Claims Tribunal cases, and even the Speech Transcription System,[7] a voice recognition system for instant transcription of court proceedings.

2 The final frontier of the use of technology in courts then appears to be the actual rendering of judicial decisions itself, although few countries appear prepared to replace judges with AI. Even China, which has been devoting significant resources into developing technologically-advanced courts and legal assistants, appears to consider the use of AI in actual judicial decision-making to be unthinkable, saying that the expertise of judges is irreplaceable, not least by AI.[8] It is of course possible to suggest several reasons for such a sentiment, such as societal confidence or lack thereof in AI, the need for the accountability that accompanies a societal presence, or the relative ease of training humans in common-sense reasoning as compared to AI. But in isolation, such a bald statement, made without substantiation, sounds merely like judicial exceptionalism, especially in light of the rapid advances that AI has made in recent years, and only exacerbates the risk of judiciaries being caught flat-footed if or when the spectre of AI judicial decision-making becomes reality.

---

wps/wcm/connect/7898d957-69b5-4727-9226-277e8ae28711/EMCompass-Note-71-AI-Investment-Trends.pdf?MOD=AJPERES&CVID=mR5Jvd6> (accessed 20 April 2020).

3 Daniel Faggella, "AI in Law and Legal Practice – A Comprehensive View of 35 Applications" *Emerj* (14 March 2020) <https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/> (accessed 20 April 2020).

4 See State Courts Singapore, "Integrated Case Management System" <https://www.statecourts.gov.sg/cws/CriminalCase/Pages/The-ICMS-portal.aspx> (accessed 20 April 2020).

5 See Community Justice Centre, "What is Automated Court Documents Assembly?" <https://cjc.org.sg/automated-court-documents-assembly/self-help-bankruptcy/frequently-asked-questions/> (accessed 20 April 2020).

6 See State Courts Singapore, "Community Justice and Tribunals System" <https://www.statecourts.gov.sg/CJTS/#!/index1> (accessed 20 April 2020).

7 Tan Tam Mei, "State Courts to Use System That Instantly Transcribes Court Proceedings" *The Straits Times* (14 December 2017).

8 Yan Jie, "China's Courts Look to AI for Smarter Judgments" *Sixth Tone* (18 November 2016) <http://www.sixthtone.com/news/1584/china%20s-courts-look-ai-smarter-judgments> (accessed 20 April 2020).

3 Whether or not AI judicial decision-making comes to pass, the pace of AI-driven change in the legal sector suggests that it may be prudent for judiciaries to begin considering now what AI judicial decision-making may look like. There is already a need for judiciaries to begin grappling with the use of AI decision-making in related fields, such as the use of algorithms in administrative decision-making.[9] Even if AI judicial decision-making fails to materialise, this exercise in thinking about what truly drives judicial decision-making may provide valuable lessons and thought leadership for a legal sector that increasingly augments its services with AI. And if AI judicial decision-making does indeed become reality, judiciaries will have drawer plans that they can rely on in taking charge of shaping the use of AI in the courts. It behoves them to ensure that courts continue to deliver the justice that society needs, even if not the justice that society sometimes may expect, and if AI were one day to make judicial decisions, judges must be there every step of the way to make sure that justice is not just done but is seen to be done.

4 This article is split into three parts. The first part deals with the current state of AI as applicable to the legal sector and developments in the field of AI. As is the case with every dynamic field, it will be quite difficult to be comprehensive, and this article will not purport to be so. Nonetheless, a taster in the form of general principles will suffice for present purposes. The second part breaks down what judicial decision-making entails, as far as is possible for laypersons to do so, in the hope of distilling the essence of judicial decision-making that any AI judicial decision-making solution must be able to replicate. The third part draws upon the analyses of the former two parts to consider what implementations of AI will satisfy the requirements of judicial decision-making, and therefore what is required to put together such an implementation of AI. It also considers if any aspects of judicial decision-making could change in response to the use of AI in judicial decision-making.

5 A key caveat is that the observations in the rest of this article on the usability of AI in judicial decision-making below hold up only for current implementations of AI as weak AI.[10] Paradigm changes in the field may invalidate these observations. For example, quantum-[11]

9 See, *eg*, Ashley Deeks, "The Judicial Demand for XAI" (2019) 119 Colum Law Rev 1829 at 1838–1840; and Monika Zalnieriute, Lyria Bennett Moses & George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82(3) Mod L Rev 425 at 427.

10 John Searle, "Minds, Brains, and Programs" (1980) 3(3) *Behavioral and Brain Sciences* 417 at 2 <http://cogprints.org/7150/1/10.1.1.83.5248.pdf> (accessed 23 July 2020).

11 See, *eg*, Nick Easen, "When Quantum Computing and AI Collide" *Raconteur* (28 April 2020) <https://www.raconteur.net/technology/quantum-computing-ai> (accessed 23 July 2020).

and photon-based[12] intelligence technologies both have the potential to significantly increase the efficiency of AI algorithms. Without practical data as to the effects of these advances in technology, it will profit us little to speculate on possible changes; being aware that such a possibility exists suffices at this point.

II. The workings of AI

A. Overview

6 AI generally refers to the phenomenon of computers "learning" to perform a task better, and hence simulating the human capacity to learn and process information. The foundations of AI research lay in the Dartmouth Summer Research Project on Artificial Intelligence in 1956, a gathering of a group of scientists interested in the field of AI, from whence came the Dartmouth Proposal: "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it".[13] Currently workable AI applications are limited to replicating specific features of human intelligence, and are known generally as "weak" AI.[14] This is in contrast to the type of AI more commonly seen in science fiction and known technically as "strong" AI – an all-round AI capable of human-like thought and interactions with other humans, but which is not currently achievable with the present state of the art.[15]

7 This pursuit of the replication of human intelligence is by no means monolithic. AI research centres around various techniques that show promise for replicating intelligence, and the viability of these techniques have waxed and waned depending on a multitude of factors, such as scientific interest, developments in allied mathematical fields, and technological advancements. Research into an early form of artificial neuron, known as the perceptron, stalled after it was argued that existing computers lacked the necessary processing power to handle

12 See, *eg*, American Institute of Physics, "Photon-based Processing Units Enable More Complex Machine Learning" *TechXplore* (21 July 2020) <https://techxplore.com/news/2020-07-photon-based-enable-complex-machine.html> (accessed 23 July 2020).

13 John McCarthy *et al*, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" (31 August 1955) at p 1, accessible at <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf> (accessed 20 April 2020).

14 John Searle, "Minds, Brains, and Programs" (1980) 3(3) *Behavioral and Brain Sciences* 417 at 2 <http://cogprints.org/7150/1/10.1.1.83.5248.pdf> (accessed 23 July 2020).

15 John Searle, "Minds, Brains, and Programs" (1980) 3(3) *Behavioral and Brain Sciences* 417 at 2 <http://cogprints.org/7150/1/10.1.1.83.5248.pdf> (accessed 23 July 2020).

large networks of perceptrons.[16] Attention then shifted to the building of expert systems, aiming to encode expert knowledge.[17] Although expert systems were very successful when they worked, the sheer complexity of the rules that governed human knowledge began to pose a problem for the updating and maintenance of such systems, leading to their gradual abandonment in favour of modern techniques.[18]

8 Current AI research revolves around machine learning – the training of an algorithm with correlated sets of input-output data, in order to predict an output given a particular input.[19] This is achieved through a range of statistical techniques such as linear and logistic regression,[20] tree models,[21] and artificial neural networks,[22] building on the back of advancements in computing and digital storage technology. In the legal context, an AI could for example be trained to predict the outcome of the division of matrimonial assets by analysing the parties' actuarial data.[23] At the point of optimisation, the AI is in theory capable of returning or predicting what the outcome of any given input that is within the parameters of the training data would be. The value of such an AI is in its ability to process far more information than humans will ever be able to. However, because such techniques adapt to large amounts of provided training data iteratively, their exact method of operation and the way in which they arrive at results is not immediately explainable, if

16 Mikel Olazaran, "A Sociological Study of the Official History of the Perceptrons Controversy" (1996) 26(3) *Social Studies of Science* 611 at 623–627.

17 See generally Wikipedia, "Expert System" <https://en.wikipedia.org/wiki/Expert_system> (accessed 23 July 2020).

18 Juri Yanase & Evangelos Triantaphyllou, "A Systematic Survey of Computer-aided Diagnosis in Medicine: Past and Present Developments" (2019) 138 *Expert Systems with Applications* 112821 at 4–5.

19 See generally Wikipedia, "Machine Learning" <https://en.wikipedia.org/wiki/Machine_learning> (accessed 20 April 2020). Machine learning is generally defined as the study of computer algorithms that improve automatically through experience.

20 Rohith Gandhi, "Introduction to Machine Learning Algorithms: Linear Regression" *Towards Data Science* (27 May 2018) <https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a> (accessed 23 July 2020).

21 George Seif, "A Guide to Decision Trees for Machine Learning and Data Science" *Towards Data Science* (30 November 2018) <https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956> (accessed 23 July 2020).

22 Victor Zhou, "Machine Learning for Beginners: An Introduction to Neural Networks" *Towards Data Science* (6 March 2019) <https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9> (accessed 23 July 2020).

23 See, *eg*, Fabian Koh, "NUS Law and Economics Student, Along with Three Peers, Creates Case Outcome Simulator" *The Straits Times* (7 January 2018).

at all,[24] although the degree to which this is true varies from technique to technique and implementation to implementation. For instance, simple decision trees, being essentially a cascading series of classifications, are relatively easy to understand, whereas the most complex form of neural networks, with multiple factors feeding into each other at different levels and degrees, can be very difficult to plot and visualise in a graphical manner.

9 The difficulty in explaining their results notwithstanding, these sophisticated statistical techniques are powerful and accurate enough that they have been experimentally and sometimes even commercially applied in situations where they are capable of delivering results that can exceed those of trained professionals. An example is the use of AI in radiology and imaging diagnosis,[25] where the performance of AI algorithms was shown to exceed that of practising radiologists in the detection of pneumonia, with the promise of extending the same techniques to the detection of other diseases. Closer to legal home, an experiment showed that one AI contract review software beat lawyers for both time and accuracy in reviewing non-disclosure agreements, albeit in looking out for specific clauses or issues that the AI was trained for.[26] While much work still needs to be done to improve the rigour of these applications for frontline deployment, these experimental results show sufficient promise for industry stakeholders to consider pursuing them in earnest, as evidenced by the surge in interest in AI applications in modern industry.[27]

10 The success of these AI techniques also owes much to the extremely modern phenomenon of "big data".[28] Human history has never before witnessed the capability not just to granularise so many aspects

24 See, *eg*, Ariel Bleicher, "Demystifying the Black Box that is AI" *Scientific American* (9 August 2017) <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/> (accessed 23 July 2020).

25 Pranav Rajpurkar *et al*, "CheXNet: Radiologist-level Pneumonia Detection on Chest X-Rays with Deep Learning" (2017) <https://arxiv.org/abs/1711.05225> (accessed on 20 April 2020).

26 See, *eg*, Alfred Chua, "The Big Read: Rise of the Machine – How Technology is Disrupting Singapore's Law Firms" *Today* (19 January 2019).

27 See, *eg*, Xiaomin Mou, "Artificial Intelligence: Investment Trends and Selected Industry Uses" (EMCompass Note 71, September 2019) <https://www.ifc.org/wps/wcm/connect/7898d957-69b5-4727-9226-277e8ae28711/EMCompass-Note-71-AI-Investment-Trends.pdf?MOD=AJPERES&CVID=mR5Jvd6> (accessed on 20 April 2020).

28 There is no official definition of "big data" especially as its applications grow ever more voluminous, although it is generally agreed that "big data" is any data set that is beyond the ability of common software to manage and evaluate in reasonable amounts of time – see generally Wikipedia, "Big Data" <https://en.wikipedia.org/wiki/Big_data> (accessed 20 April 2020).

of human existence, but also to analyse and utilise such massive flows of information. The use of big data permits AI to rapidly analyse trends within the data itself, in a way not normally understandable by humans, and in so doing aim to return a result that approximates what humans would have arrived at if they were capable of parsing and understanding all that data in a reasonable time with traditional software.[29] In the legal context, big data can refer to a variety of materials such as case law,[30] standard templates, and even – in the context of copyright complaints under the Digital Millennium Copyright Act[31] – form-based notifications submitted to content hosting intermediaries.[32] Lawyers have developed various strategies to index and isolate materials in the above categories that fall within their field of interest such as modern commercial databases, but it is near impossible for any one person to have an overview, much less mastery, of all the information contained in modern legal databases in the same way that AI can.

11 An additional complexity introduced by legal big data is the fact that the vast majority of legal materials are in the form of text, requiring intermediary steps in the form of natural language processing ("NLP") before the information contained within can be usefully parsed by AI.[33] Modern NLP relies on statistical processing of existing texts, or *corpora*, in order to attempt to deduce the uses of words based on the context of their occurrences.[34] It is generally accepted that computers do not quite understand language in the same way that humans do, given that computers derive their understanding of language endogenously from trends within the texts themselves, whereas humans use language to mediate and share their understanding of the real world with others.[35] Nonetheless, AI NLP applications have been able to approximate to

29 See generally Wikipedia, "Big Data (Definitions)" <https://en.wikipedia.org/wiki/Big_data#Definition> (accessed 23 July 2020).

30 See, *eg*, Bernard Marr, "How Big Data is Disrupting Law Firms and the Legal Profession" *Forbes* (20 January 2016).

31 Pub L No 105-304, 112 Stat 2860 (28 October 1998) (US).

32 See, *eg*, Daniel Seng, "The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices" (2014) 18 Va JL & Tech 369 at 378–383.

33 See generally Wikipedia, "Natural Language Processing" <https://en.wikipedia.org/wiki/Natural_language_processing> (accessed 23 July 2020). While it is conceivably possible for smaller scale applications to have humans read and translate legal text into data points that can be processed by traditional mathematics, natural language processing techniques are required when dealing with larger *corpora* of legal data that would be impractical for humans to process in a reasonable amount of time.

34 See generally Wikipedia, "Natural Language Processing" <https://en.wikipedia.org/wiki/Natural_language_processing> (accessed 23 July 2020).

35 See, *eg*, Karen Hao, "AI Still Doesn't Have the Common Sense to Understand Human Language" *MIT Technology Review* (31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> (accessed 23 July 2020).

a certain degree what we would consider a rudimentary understanding of language, and certainly to an extent that is prototypically usable in the legal sector.[36]

B. Black boxes and xAI

12 User trust in the reliability of AI-generated results is a major problem, especially in the legal context, for four reasons which flow immediately from the above premises. First, the operation of modern AI algorithms depends on intricate iterative mathematical operations which the average layperson cannot easily follow.[37] Second, such operations rely on large datasets which by definition are difficult to parse in manageable or productive time frames with traditional software.[38] Third, much of the data that the legal sector possesses requires the intervention of NLP techniques to convert into big data suitable for AI use, which introduces an additional layer of uncertainty and error due to the very different ways in which NLP techniques and humans derive meaning from text.[39] Lastly, legal results derive their legitimacy from the validity of the reasoning process used to derive said result – if the reasoning process cannot be explicated, the result cannot be defended. The importance of legal due process to legal decision-making is demonstrated by the various fundamental rules of natural justice that underpin our legal system, and to a lesser extent, the Rules of Court that regulate legal proceedings in the interest of guaranteeing that both parties can have their day in court fairly and justly.

13 These problems are by no means unique to the legal sector. The sheer scale and opacity of big data and the AI techniques that utilise it has led to suggestions that as long as AI results match or exceed that

36 See, *eg*, Richard Tromans, *Legal AI – A Beginner's Guide* (Thomson Reuters) (20 February 2017) at p 3 <https://blogs.thomsonreuters.com/legal-uk/wp-content/uploads/sites/14/2017/02/Legal-AI-a-beginners-guide-web.pdf> (accessed 23 July 2020).

37 See, *eg*, David Weinberger, "Our Machines Now Have Knowledge We'll Never Understand" *Wired* (4 October 2017) <https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/> (accessed 23 July 2020).

38 See generally Wikipedia, "Big Data (Definitions)" <https://en.wikipedia.org/wiki/Big_data#Definition> (accessed 23 July 2020).

39 See, *eg*, Karen Hao, "AI Still Doesn't Have the Common Sense to Understand Human Language" *MIT Technology Review* (31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> (accessed 23 July 2020); and Jesse Dunietz, "The Field of Natural Language Processing is Chasing the Wrong Goal" *MIT Technology Review* (31 July 2020) <https://www.technologyreview.com/2020/07/31/1005876/natural-language-processing-evaluation-ai-opinion/?itm_source=parsely-api> (accessed 31 July 2020).

of professionals, it matters little that humans do not know how the results were derived – an approach to AI known as the "black box".[40] The acceptability of the "black box" approach in any given context depends on factors including but not limited to the importance of the AI's decision, the legitimacy of a decision independent of its chain of reasoning, and any biases or preferences the AI exhibits or has been programmed to exhibit.[41]

14 While the black box approach may suffice for relatively trivial matters, the lack of understanding as to how the results were derived is wholly unsatisfactory for any application where the decision-making process has legal ramifications in so far as that process affects the rights of individuals.[42] Without knowing how the AI arrives at decisions, the decisions it issues can justifiably be called arbitrary.

15 Here, executive decision-making and judicial decision-making need to be distinguished by the degree of transparency inherent in the issued decision. In general, judicial decision-making is more transparent than executive decision-making due to the judicial duty to give reasons,[43] and by extension the fundamental rules of natural justice that underpin the workings of any legal system. In contrast, executive decision-making affords the decision-maker a degree of discretion, generally challengeable only by the significant expense of judicial review (which is generally restricted to the process and not the merits of the finding),[44] and then even some statutes provide that judicial review shall not lie in certain matters.[45] As the use of AI in executive decision-making grows, whether

40 See, *eg*, Ariel Bleicher, "Demystifying the Black Box that is AI" *Scientific American* (9 August 2017) <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/> (accessed 23 July 2020). For instance, although we may not fully understand the exact biological operating mechanisms of drugs, this does not prevent them from being used to treat critical conditions, if the regulators are satisfied as to their safety record in clinical trials.

41 See, *eg*, Will Knight, "The Dark Secret at the Heart of AI" *MIT Technology Review* (11 April 2017) <https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/> (accessed 23 July 2020). There is the possibility that our own human intelligence is also incompletely explainable, and it would therefore be futile to expect humans to be able to build a completely explicable artificial intelligence.

42 See, *eg*, Ashley Deeks, "The Judicial Demand for XAI" (2019) 119 Colum L Rev 1829 at 1843–1845; and Monika Zalnieriute, Lyria Bennett Moses & George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82(3) Mod L Rev 425 at 428.

43 See Supreme Court of Singapore, *Judicial Code of Conduct for the Judges and Judicial Commissioners of the Supreme Court of Singapore* (20 February 2019) at pp 13–14; and *Thong Ah Fat v Public Prosecutor* [2012] 1 SLR 676.

44 *Chee Siok Chin v Public Prosecutor* [2006] 1 SLR(R) 582 at [93].

45 See, *eg*, Internal Security Act (Cap 143, 1985 Rev Ed) s 8B(2).

by way of black box techniques or otherwise, the courts themselves will need to be *au fait* with AI, and to develop their own strategies to examine the use of AI by the Executive. In the latter case, courts will still need to at least explicate the relevant statute for the benefit of the present parties and future litigants.

16 If we insist, as we should, upon the continued existence of the judicial duty to give reasons and the fundamental rules of natural justice as the foundations of our legal system, we must categorically reject the black box approach to AI for the purposes of AI judicial decision-making, and expect a minimum degree of explanation from AI. This is especially if, in their traditional role as a check-and-balance on the exercise of legislative and executive power in the separation of powers model of governance, the Judiciary is expected to examine the decision-making process of the Executive which may involve the use of AI. It would be inconceivable for the Judiciary themselves to use black box algorithms to pass judgment on the use of black box algorithms by the Executive.[46]

17 Thankfully, there is increasing interest and emphasis in designing AI according to human-centric design principles. The increasing profile of AI has led to a corresponding increase in interest in its governance, design and ethics. Academics and AI developers themselves have argued for the need to be accountable and transparent by design, and there are multiple public and private efforts to develop a code of ethics for the use of AI, but this requirement for transparency has yet to gain traction as a formal requirement imposed by states.[47] Various major organisations have convened their own AI advisory boards or released their own AI standards, such as Microsoft's Aether[48] committee and the Personal Data Protection Commission's *Model AI Governance Framework*, now into its second edition.[49] Such efforts in establishing some form of standards for the use of AI concur that the design of AI must have at its core human

46 Ashley Deeks, "The Judicial Demand for XAI" (2019) 119 Colum L Rev 1829 at 1841–1842.

47 European Parliament, *Civil Law Rules on Robotics* (P8_TA(2017)0051, 16 February 2017) <https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html> (accessed 20 April 2020). The EU Parliament Legal Affairs Committee has recommended the values of inclusiveness and accountability to be part of the ethos of robotics engineers, which the EU Parliament has formally adopted but which has yet to manifest itself in the form of any specific guidelines or directives.

48 "Aether" stands for artificial intelligence ("AI"), Ethics and Effects in Engineering and Research, and is an advisory board at Microsoft focusing on potential issues related to developing and fielding AI applications. See Matt Weinberger, "Microsoft is Forming a Grand Army of Experts in the Artificial Intelligence Wars with Google, Facebook, and Amazon" *Business Insider* (12 July 2017).

49 Personal Data Protection Commission, *Model AI Governance Framework* (2nd Ed, 21 January 2020).

well-being and safety, and must be explainable, transparent and fair,[50] and this extends to the incorporation of these values in regulations and laws concerning AI.[51] This has led to renewed and increased research interest in "explainable AI" ("xAI"), which seeks to remove or ameliorate the unknown quality from AI decision-making, and to be able to provide human-comprehensible reasons for the decisions that an AI makes.

18 xAI approaches can be categorised into two main types: exogenous and decompositive.[52] Exogenous xAI approaches can be further divided into model-centric explanations or subject-centric explanations. Model-centric explanations focus on providing broad information on how an AI works without dealing with individual examples, such as providing information on how an AI was programmed, the parameters of the training data fed into it, and performance metrics such as its accuracy, precision and recall on test data.[53] Subject-centric explanations focus instead on explaining a particular output by comparing or contrasting it with other outputs, providing a sense of how changes in the input attributes of a particular case affect its output classification.[54] Decompositive xAI approaches instead attempt to explain the very operation of an AI itself, whether by bluntly exposing its source code[55] or by constructing another model that seeks to reconstruct the correlation between inputs and outputs via repeated queries, and cast some illumination on the working of the original model in the process.[56]

---

50 Personal Data Protection Commission, *Model AI Governance Framework* (2nd Ed, 21 January 2020) at pp 15 and 64–66.

51 Singapore Academy of Law, Law Reform Committee, *Applying Ethical Principles for Artificial Intelligence in Regulatory Reform* (July 2020) at para 1.9.

52 In the interests of ease of reference and some degree of standardisation of terminology, this article will adopt the same terminology as used by Ashley Deeks, "The Judicial Demand for XAI" (2019) 119 Colum L Rev 1829 at 1835, but will similarly adopt the caveat at fn 28 that the "discussion of categories of xAI is necessarily simplified, because there are a wide range of approaches to categorizing xAI and the nomenclature is unsettled". See also Daniel Seng & Stephen Mason, "Artificial Intelligence and Evidence" (2021) 33 SAcLJ 241 at 277, para 66.

53 Lilian Edwards & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16(1) Duke L & Tech Rev 18 at 55–56.

54 Lilian Edwards & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16(1) Duke L & Tech Rev 18 at 57–58.

55 Lilian Edwards & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16(1) Duke L & Tech Rev 18 at 64–65.

56 Lilian Edwards & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16(1) Duke L & Tech Rev 18 at 64–65.

19 Designing xAI unfortunately necessarily requires much more effort than designing AI *simpliciter*. Part of the appeal of modern AI approaches lies in their simplicity and democratisation relative to their power. In contrast to the past "expert systems" approaches to AI which often entailed painstaking hand-coding of rules and hence mandated the commitment of significant expertise and resources,[57] the entry requirements for developing modern AI systems are much lower and more amenable, especially with the widespread availability of datasets and open-source AI libraries.[58] The corollary of this ease and power is the abstraction of the core algorithms within such an AI, hiding its fundamental workings from the user.[59] Regaining that knowledge, whether by way of exogenous or decompositive xAI, in effect renders moot the ease of development of modern AI since effort must now also be put into penetrating the very abstraction that made AI accessible in the first place. In further contrast to expert systems, which were human-comprehensible from the start by virtue of being if-then encodings of human reasoning,[60] more effort is required to translate an AI's mathematical and statistical reasoning to a form that humans can understand. In addition, some argue for the black box approach on the ground that depending on the exact designers and implementation of xAI, there may be legal and economic ramifications such as a Hobson's choice between innovating and being forced to reveal trade secrets regarding the operation of the AI, or not innovating at all.[61] While this is inaccurate in so far as these trade-offs can be mitigated by frameworks mirroring the existing intellectual property regimes for incentivising research and development, or by other means of regulation, oversight or validation,[62] the fact remains that the development of xAI still requires much more work than the mere implementation of AI for the reasons above.

---

57 Philip Leith, "The Rise and Fall of the Legal Expert System" (2010) 1(1) EJLT 1.

58 See, *eg*, Gordon Haff, "How to Get Started in AI" *opensource.com* (10 December 2018) <https://opensource.com/article/18/12/how-get-started-ai> (accessed 20 April 2020).

59 See, *eg*, Justin Gage, "Machine Learning Abstraction and the Age of AI Ease" *Machine Learnings* (14 September 2017) <https://machinelearnings.co/machine-learning-abstraction-and-the-age-of-ai-ease-f3274bb9e0c1> (accessed 23 July 2020).

60 Lilian Edwards & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You Are Looking For" (2017) 16(1) Duke L & Tech Rev 18 at 64–65.

61 See Ashley Deeks, "The Judicial Demand for XAI" (2019) 119 Colum L Rev 1829 at 1834; and Finale Doshi-Velez *et al*, "Accountability of AI under the Law: The Role of Explanation" (2017) at p 2 <https://arxiv.org/ftp/arxiv/papers/1711/1711.01134.pdf> (accessed 20 April 2020). Such costs are more apparent in decompositive xAI, although even exogenous xAI may allow others to deduce or reverse engineer more aspects of the AI than its developers would be comfortable with.

62 See, *eg*, Daniel Seng & Stephen Mason, "Artificial Intelligence and Evidence" (2021) 33 SAcLJ 241 at 272–276, paras 60–64.

20 Costs notwithstanding, if AI were to be used at all by judiciaries, then the specific use of xAI by judiciaries is non-negotiable. It has already been shown above that the use of black box approaches to AI involved in judicial decision-making is by definition unacceptable. If we accept that xAI is the only form of AI that is acceptable for use by judiciaries in AI judicial decision-making, the question then is what form that xAI will take and how it ensures the continued confidence of society in the legal system. The answer depends on what society expects from judicial decision-making, which we now turn to examine.

III. Judicial power, functions and duties

A. Judicial power and functions

21 The Judiciary as a whole is the executor of the judicial power vested in the Supreme and State Courts by virtue of Art 93 of the Constitution of the Republic of Singapore.[63] While difficult (or perhaps impossible) to define exhaustively, the judicial power is generally agreed to be the power that "every sovereign authority must of necessity have to decide controversies between its subjects or between itself and its subjects, whether the rights relate to life, liberty or property", and the judicial function to entail "the courts making a finding on the facts as they stand, applying the relevant law to those facts and determining the rights and obligations of the parties concerned for the purposes of governing their relationship for the future".[64]

22 From those quotations by Chan Sek Keong CJ, a number of elements of the judicial function are immediately evident:

(a) the making of findings of fact;

(b) the application of the law to said findings of fact; and

(c) determining the rights and obligations of the parties concerned.[65]

63 1999 Reprint.

64 *Mohamed Faizal bin Sabtu v Public Prosecutor* [2012] 4 SLR 947 at [20]–[21], *per* Chan Sek Keong CJ, citing Griffith CJ in *Huddart, Parker and Co Pty Ltd v Moorehead* (1909) 8 CLR 330 at 357 and Kitto J in *The Queen v The Trade Practices Tribunal, ex parte Tasmanian Breweries Pty Ltd* (1970) 123 CLR 361 at 374–375.

65 For completeness, the judicial function arguably also includes other ancillary powers that the court necessarily must have for the proper exercise of its judicial power, such as the power and the discretion to regulate its own procedure. However, as the core interest of this article lies in providing a framework for the compatibility of judicial decision-making with xAI, these ancillary powers will not be further dealt with here.

(1) Making of findings of fact

23 In their dual role as both triers of fact and the law, judges must decide, based on the evidence adduced by the parties, whether certain legally significant facts are made out or not. This is in effect an attempt to reconstruct the series of events giving rise to the instant case to the highest possible fidelity with the assistance of the evidence of the parties concerned, although the court must do so while relying on inferior historical accounts from witnesses, documents and articles.[66] Logically, only when the facts have been ascertained (or otherwise) can the law be applied to the facts. In practice, the legal cart may well be put before the factual horse, such as in cases where parties are *ad idem* very early on as to the application of a legal rule and only dispute whether the underlying facts are or are not true.[67] The finding of fact will then support the legal elements of the pleadings, and by extension, the outcome of the case.[68] Conceptually, this process may be repeated as many times as necessary for parties to satisfy their burden of proof – for example, a plaintiff may need to adduce evidence to prove first that a binding contract was made, and then adduce evidence to prove a breach of the same.

24 It would naturally be remiss to discuss how judges make findings of fact without discussing the law of evidence. Judges must make findings of fact based on the evidence that parties lead,[69] which is in turn governed by the law of evidence that aims to qualitatively maximise the chances of arriving at a safe verdict.[70] Put quasi-mathematically, the law of evidence has as its ultimate aim the maximising of true positives and negatives, or in other words, ensuring that the legal finding for or against a party most often, if not always, matches the facts as they happened.

25 At the most basic level, fact-finding – whether in civil or criminal cases – is best understood as a two-stage process. First, the evidence of parties is adduced, subject to the laws of evidence that govern the litigation process. Second, based on the evidence adduced, the judge determines which party's facts are better supported on the evidence – or

66 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 1.006.

67 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 16.009.

68 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 16.006.

69 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 1.034.

70 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at paras 1.003–1.006.

even, if a party has failed to meet the burden of proof, unsupported at all.[71]

26 To do so, judges are assisted by the definitions of "proved", "disproved" and "not proved" in the Evidence Act.[72] The definition of "proved" in the Evidence Act requires that the court, after considering the matters before it, believes that a fact either exists or that its existence is so probable that a prudent man ought, in the circumstances of the particular case, to act upon the supposition that it exists. The reference to the circumstances of the particular case refers to the standard of proof required for the case, in other words the standards of the balance of probabilities and beyond a reasonable doubt in civil and criminal cases respectively.[73]

27 While pithy to cite, these standards of proof are not easy to define in mathematical terms, and even harder to use in a mathematical sense in the process of legal reasoning. Take, for instance, the definition of the balance of probabilities. Quantitatively, it is perhaps easy to say that a balance of probabilities is a mathematical, better than 50% chance of a particular case being correct or plausible.[74]

28 However, actual attempts to utilise this mathematical formulation in legal reasoning have come to grief, the English case of *Re A (children)*[75] being most instructive in this regard. At first instance, the judge drew on the judgment of Mostyn J in *A County Council v M & F*[76] for the proposition that where there were rival hypotheses put forth by the parties, the judge was not bound to make a finding if, in his opinion, neither of the hypotheses rose to the burden of proof.[77] He then proceeded to assign a rough probability to the three possible explanations for the facts, pointing out that his aggregated figures showed that none of the explanations rose to the burden of proof applicable.[78]

29 However, King LJ disapproved of the trial judge's summation of probabilities, stating that she "[could not] agree that the use of percentages

71 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at paras 1.003–1.006.
72 Evidence Act (Cap 97, 1997 Rev Ed) ss 2(3)–2(5).
73 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 12.001.
74 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 12.091.
75 [2018] EWCA Civ 1718.
76 [2012] 2 FLR 939; [2011] EWHC 1804 (Fam).
77 *Re L (a child)* [2017] EWHC 3707 (Fam) at [98].
78 *Re L (a child)* [2017] EWHC 3707 (Fam) at [98].

and or 'aggregation' [was] the proper approach to the judicial function in respect of the simple application of the balance of probabilities".[79] King LJ went further to disapprove of Mostyn J's judgment in *A County Council v M & F*, holding that his approach there was not the appropriate approach to the balance of probabilities.[80] Instead, King LJ chose to draw upon the judgments of Lord Brandon in *Fenton Insurance Co Ltd v Rhesa Shipping Co SA (The Popi M)*[81] and Toulson LJ in *Nulty Deceased v Milton Keynes Borough Council*[82] to imply that attempts to assign probabilities to hypotheses was overly formulaic and intrinsically unsound, and to explicitly hold that judgment was not to be made by reference to percentage possibilities or probabilities.[83]

30 Against this withering attack on probabilistic reasoning, Mostyn J, in collaboration with a statistician, attempted to mount a defence by arguing that there was still room for the use of probabilistic reasoning in the present legal orthodoxy of what he termed "inference to the best explanation".[84] But even he had to concede that the role of probabilistic reasoning in judicial decision-making was limited to its use as statements of subjective degrees of belief[85] and the explication, rather than the replacement, of the fact-finding process.[86]

31 If the standard of balance of probabilities, despite its quasi-probabilistic cant, is difficult to mathematically reconcile, the standard of beyond reasonable doubt is outright impossible to quantitatively define. The Singapore courts have noted that it is unprofitable, circular even, to attempt to define the standard of beyond reasonable doubt in quantitative terms,[87] preferring instead the qualitative definition of a "reasoned doubt supported by the evidence". The emphasis is therefore placed, not on the likelihood of the circumstances of the accused's innocence, but on the cogency of the evidence in proving said circumstances.

32 For all the preference of the courts for non-qualitative measurements of the burden of proof, probabilistic reasoning itself still

79 *Re A (children)* [2018] EWCA Civ 1718 at [51].
80 *Re A (children)* [2018] EWCA Civ 1718 at [54].
81 [1985] 1 WLR 948.
82 [2013] 1 WLR 1183; [2013] EWCA Civ 15.
83 *Re A (children)* [2018] EWCA Civ 1718 at [58].
84 Ian Hunt & Mostyn J, "Probability Reasoning in Judicial Fact-finding" (2020) 24(1) *The International Journal of Evidence & Proof* 75 at 84–85.
85 Ian Hunt & Mostyn J, "Probability Reasoning in Judicial Fact-finding" (2020) 24(1) *The International Journal of Evidence & Proof* 75 at 84–85.
86 Ian Hunt & Mostyn J, "Probability Reasoning in Judicial Fact-finding" (2020) 24(1) *The International Journal of Evidence & Proof* 75 at 84–85.
87 *Public Prosecutor v GCK* [2020] 1 SLR 486 at [126]–[128], citing V K Rajah JA in *Jagatheesan s/o Krishnasamy v Public Prosecutor* [2006] 4 SLR(R) 45 at [55].

has important evidential roles to play in support of the law. It is relatively uncontroversial that fingerprinting, DNA and blood testing evidence have developed to a degree as to make them relatively indispensable in forensics and proof of identity when correctly used and explained,[88] despite some missteps along the way.[89] But the experience of the English courts shows that it is one thing to accept probabilistic reasoning in the evaluation of one aspect of the evidence, and another to use probabilistic reasoning in an overall assessment of the evidence. It remains to be seen whether there will be occasion to review the English courts' strong disapproval of probabilistic reasoning in overall fact-finding in the Singapore courts.

(2) Applying the law

33 Once the facts have been determined, the judge must then determine the law to be applied. In every case, the applicable law is always a matter of interpretation, and the question is rather how simple or complex that interpretation is depending on both the source of the law and the particular fact pattern of a case itself.

34 Relatively simple cases may involve the interpretation of straightforward statutory provisions in a paradigm case well within the conception of lawmakers when that particular statute was passed, with no unusual features such as aggravating or mitigating factors. Even in such cases, especially the first few test cases, the parties involved will usually have to have recourse to Hansard to confirm the intention of Parliament.[90] More complicated cases may result from cases falling to be decided by judge-made law, but with fact patterns and policy considerations distant

88 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 8.013.

89 See, *eg*, *R v Clark* [2003] EWCA Crim 1020 at [94]–[110] and [172]–[180]. Sally Clark was convicted of killing her children, in part after expert evidence that quoted the risk of two sudden infant death syndrome cases in the same family as one in 73 million, and as likely as "long odds winners of the Grand National year after year". The statistic was later argued and accepted to be oversimplistic. In fairness, the trial judge recognised the danger of relying on statistics and attempted to warn the jury accordingly (albeit by baldly telling them that "we do not convict people in these courts on statistics" – a statement that wholly overlooks the nature of modern forensic evidence), and the second appellate court opined that the first appellate court would have allowed the appeal due to this misuse of statistical evidence if not for the remainder of the evidence. The second appeal succeeded only because it was discovered that scientific evidence suggesting an otherwise natural death from infection had not been led at the earlier appeals, and not on the basis that there had been a mistrial by the oversimplified use of the one in 73 million statistic.

90 See generally s 9A of the Interpretation Act (Cap 1, 2002 Rev Ed). See also *Public Prosecutor v Lee Ngin Kiat* [1992] 3 SLR(R) 955 at [29] for an example of how this may be done.

from those upon which older decisions were made. On relatively rare occasions, the socio-legal circumstances underlying those decisions may have changed or shifted as well, giving judges some serious food for thought in determining whether or not the existing law ought to be changed before determining the rights and obligations of the parties concerned.[91] Therefore, what a court is judging in those cases are not facts *per se*, but facts in relation to the policy espoused in earlier cases – a task which requires an understanding and appreciation of the policy behind those cases.

(3) Determining the rights and obligations of parties concerned

35 Even after discerning the appropriate law applicable to a case, the devil may remain in the details of determining the rights and obligations of the parties concerned. A good example of such details can be found in the field of criminal sentencing. It is true that trends are generally observable in criminal sentencing and that aggravating and mitigating factors in criminal sentencing can be isolated as having some impact on the sentence. However, it is the consideration of what constitutes an aggravating or mitigating factor, and to what degree alike or different from prior cases, which poses the most difficulty for a judge; this part of the sentencing discretion is an exercise not easily reducible to mathematical formulae. Furthermore, despite an overt emphasis on equality before the law, it is also arguable that the legal system is not overly concerned with translating this equality into precise mathematical exactitude, given that the threshold for appellate intervention in sentencing is that the sentence must be "manifestly" excessive or insufficient, and not merely excessive or insufficient.[92]

36 Recent cases indicate that while mathematical reasoning does assist legal reasoning to some degree in standardising sentencing precedents, a crucial part of its use is knowing when *not* to utilise it, and this discretion can only come from an appreciation of the legal background behind its use. Offences under the law come in a range of definitions, and approaches that commend itself to some types of offences may not be applicable to others.[93]

91 See, *eg*, *See Toh Siew Kee v Ho Ah Lam Ferrocement (Pte) Ltd* [2013] 3 SLR 284 at [52]–[54] and [98]–[100]; and *Public Prosecutor v Hue An Li* [2014] 4 SLR 661 at [103]–[104].

92 See, *eg*, *Public Prosecutor v Mohammed Liton Mohammed Syeed Mallik* [2008] 1 SLR(R) 601 at [81]–[84].

93 See, *eg*, *Ng Kean Meng Terence v Public Prosecutor* [2017] 2 SLR 449 at [25]–[38].

37 In *Vasentha d/o Joseph v Public Prosecutor*[94] ("*Vasentha*"), a case involving the sentencing of a first-time drug trafficker for possession of diamorphine for the purpose of trafficking under s 5 of the Misuse of Drugs Act,[95] Sundaresh Menon CJ engaged in a rudimentary form of reasoning from linear regression when he observed from a graph of first-time drug offenders[96] that "beyond a general upward trend, the relationship [of drugs trafficked to sentence] is less than clear".[97] He then warned that precisely because of this weak relationship, the quantity of drugs trafficked was not the be-all and end-all of sentencing; the sentencing judge had to have regard to all other relevant factors including the culpability of the offender and the presence of aggravating or mitigating factors.[98] With reference to the "full spectrum" principle of sentencing,[99] Menon CJ then laid down indicative starting points for sentencing (the *Vasentha* approach), but stressed that judges must have full regard to assessing the offender's culpability and any aggravating or mitigating factors.[100]

38 The *Vasentha* approach was subsequently approved by the Court of Appeal in *Suventher Shanmugam v Public Prosecutor*.[101] However, the major caveat to the *Vasentha* approach – that the indicative starting points were suitable only for offences which largely depended on a single metric – was not explicitly stated in *Vasentha* itself, but noted separately by the Court of Appeal in *Ng Kean Meng Terence v Public Prosecutor*[102] and *Public Prosecutor v Lai Teck Guan*.[103] Some judges correctly intuited this distinction and avoided the use of the *Vasentha* approach in unsuitable cases,[104] but others failed to do so.

39 One such case was *Public Prosecutor v Lai Teck Guan*, where Sundaresh Menon CJ criticised the court at first instance for not having

94 [2015] 5 SLR 122.
95 Cap 185, 2008 Rev Ed.
96 The use of graphs by judges in explaining sentencing decisions also has other precedent in Singapore case law. See, *eg*, *Public Prosecutor v Chow Chian Yow Joseph Brian* [2016] 2 SLR 335 at [61].
97 *Vasentha d/o Joseph v Public Prosecutor* [2015] 5 SLR 122 at [30].
98 *Vasentha d/o Joseph v Public Prosecutor* [2015] 5 SLR 122 at [34].
99 The "full spectrum" principle of sentencing holds that courts should have regard to the range of sentences passed by Parliament for a particular offence to determine where in that range the accused's conduct lies. See, *eg*, *Angliss Singapore Pte Ltd v Public Prosecutor* [2006] 4 SLR(R) 653 at [86]; *Ong Chee Eng v Public Prosecutor* [2012] 3 SLR 776 at [24]; and *Poh Boon Kiat v Public Prosecutor* [2014] 4 SLR 892 at [60].
100 *Vasentha d/o Joseph v Public Prosecutor* [2015] 5 SLR 122 at [48].
101 [2017] 2 SLR 115 at [28].
102 [2017] 2 SLR 449 at [30].
103 [2018] 5 SLR 852 at [26].
104 *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852 at [27]–[28].

given sufficient regard as to whether the *Vasentha* approach was suitable for repeat offenders. Referring to the legislation, Menon CJ observed that Parliament had clearly envisaged a difference in punishment between first-time and repeat offenders, albeit somewhat inconsistently.[105] (The punishments for first-time and repeat offenders varied only up to trafficking in 10g of diamorphine; beyond that, the punishments for both first-time and repeat offenders were the same.)[106] Menon CJ held that this inconsistency rendered the use of a simple uplift based on the *Vasentha* approach inaccurate, and required a modified sentencing framework to separately account for a repeat offender "uplift".[107]

40 The modified sentencing framework for repeat offenders in *Public Prosecutor v Lai Teck Guan* was, however, not followed by Chan Seng Onn J in *Soh Qiu Xia Katty v Public Prosecutor*[108] on the ground that this approach was not wholly mindful of the "full spectrum principle" due to gaps and changes in the gradient of the indicative starting sentences that it laid down.[109] Chan J then attempted to re-calibrate the sentencing framework in *Public Prosecutor v Lai Teck Guan*, but was then rebuked by a three-member Court of Appeal (including Menon CJ) in *Mohd Akebal s/o Ghulam Jilani v Public Prosecutor*,[110] which stated that the seeming differences in *Public Prosecutor v Lai Teck Guan* and *Soh Qiu Xia Katty v Public Prosecutor* were "matters of detail and did not yield any difference in outcome".[111] The court in *Mohd Akebal s/o Ghulam Jilani v Public Prosecutor* stated that the guidelines were only a means to the end of deriving broadly consistent sentences in broadly similar cases, that they were not intended to yield mathematically perfect graphs, and

105 *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852 at [32]–[37].
106 *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852 at [32]–[37].
107 *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852 at [42]. This discrepancy in prescribed sentences meant that the repeat offender uplift, referred to as the indicative uplift, was inversely proportionate to the quantity trafficked. However, as the starting point for the sentence was still the proportionate framework established under *Vasentha d/o Joseph v Public Prosecutor* [2015] 5 SLR 122, the net effect was that sentences for repeat offenders were still generally higher than those for first-time offenders, although the rate of change with the quantity of drug imported was lesser for repeat offenders than it was for first-time offenders.
108 *Soh Qiu Xia Katty v Public Prosecutor* [2019] 3 SLR 568. *Soh Qiu Xia Katty v Public Prosecutor* also involved a repeat offender originally sentenced under the first-time offender framework in *Vasentha d/o Joseph v Public Prosecutor* [2015] 5 SLR 122 prior to the decision in *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852; both parties agreed on appeal that *Public Prosecutor v Lai Teck Guan* was now the applicable precedent for repeat offenders.
109 *Soh Qiu Xia Katty v Public Prosecutor* [2019] 3 SLR 568 at [24]–[37].
110 [2020] 1 SLR 266.
111 *Mohd Akebal s/o Ghulam Jilani v Public Prosecutor* [2020] 1 SLR 266 at [20].

were meant to be applied as a matter of common sense. Menon CJ went on to drive home this point in two further cases.[112]

41 These cases generally show that the Singapore courts view mathematical reasoning in sentencing cases at least as a useful aid in achieving broad sentencing consistency. Beyond this, however, the exact degree to which mathematical reasoning can supplement legal reasoning is unclear. Chao Hick Tin JA in *Loo Pei Xiang Alan v Public Prosecutor* noted the similarity in sentencing ranges for first-time offenders of trafficking in diamorphine and methamphetamine, and produced an "exchange rate" of 1g of diamorphine to 16.7g of methamphetamine in order to adapt the *Vasentha* approach;[113] this was noted by Menon CJ in *Public Prosecutor v Lai Teck Guan* but otherwise passed without comment.[114] Yet a similar attempt at mathematical precision in *Soh Qiu Xia Katty v Public Prosecutor*, not entirely without legal basis, was dismissed as "matters of detail".[115] It is not immediately clear where the line lies, although it is possible that any approach that has the potential to calcify judges' sentencing discretions is to be avoided.

(4) Mathematics and the judicial function – A coda

42 The record of the English and Singapore courts in engaging with mathematical reasoning alongside legal reasoning is therefore mixed. On the one hand, courts do and continue to engage with mathematical reasoning in making findings of fact. On the other hand, the English courts have demonstrated a severe reluctance to mix legal reasoning with references to mathematics; the trial judge in *R v Clark*[116] opined that "[h]owever compelling [the jury] may find [statistics] to be, we do not convict people in these courts on statistics. It would be a terrible day if that were so", and the Court of Appeal in *Re A (children)*, as we have already seen, thought little of the use of probability in explicating legal reasoning.[117]

43 This mixed record notwithstanding, there are elements of hope for mathematical reasoning alongside legal reasoning in the Singapore courts. This early exploration of mathematically rigorous sentencing at least establishes an explicit precedent for the use of mathematics in

112 See *Public Prosecutor v Siow Kai Yuan Terence* [2020] 4 SLR 1412 at [87]; and *Mao Xuezhong v Public Prosecutor* [2020] SGHC 99 at [62]. Sundaresh Menon CJ was on the bench in both of these appeals.
113 *Loo Pei Xiang Alan v Public Prosecutor* [2015] 5 SLR 500 at [17].
114 *Public Prosecutor v Lai Teck Guan* [2018] 5 SLR 852 at [39].
115 *Mohd Akebal s/o Ghulam Jilani v Public Prosecutor* [2020] 1 SLR 266 at [20].
116 *R v Clark* [2003] EWCA Crim 1020 at [104].
117 *Re A (children)* [2018] EWCA Civ 1718 at [51]–[59].

sentencing, albeit tempered by a good deal of legal common sense. When used right, the potential for mathematics to, among other things, help set legal precedent on a firm footing that accords with the intent of Parliament is there. It is true that there have been some missteps, but missteps are always to be expected whenever trying something new; furthermore, the insights and experience gained from such attempts are invaluable in looking forward to a legal future where AI plays a bigger role.

44 Nor is the use of mathematics restricted only to judges. Nothing restricts either the Prosecution or the Defence from examining the sentencing regime applicable to repeat traffickers in diamorphine and from drawing the same conclusions about the necessary applicable uplift, had they been alive to the possibility of doing so; they would have been of greater assistance to the court in that case. In addition, anecdotally, the jokes about lawyers doing law to avoid math are gaining traction as forms of ironic self-deprecation, showing at least some awareness amongst legal stakeholders of the need to have some cognisance of mathematics in legal work. This increasing awareness is a harbinger of lawyers beginning to explore whether mathematics can be used to encode legal meaning, which in turn will lay the groundwork for a discussion about whether it is both proper and acceptable to do so. However, it remains to be seen whether there will be further efforts in this regard, perhaps fuelled by further sophistication in empirical legal analytics.

B. Judicial duties

45 It is settled under both law and judicial codes of conduct that judges are under certain duties and obligations in the execution of their judicial roles,[118] which include the duty to give reasons,[119] and the duty to ensure their own freedom from actual or apparent bias.[120] Above all else, these duties hark back to the obligation that judges have, as custodians of

118 See generally Supreme Court of Singapore, *Judicial Code of Conduct for the Judges and Judicial Commissioners of the Supreme Court of Singapore* (20 February 2019).

119 See Supreme Court of Singapore, *Judicial Code of Conduct for the Judges and Judicial Commissioners of the Supreme Court of Singapore* (20 February 2019) at pp 13–14; and *Thong Ah Fat v Public Prosecutor* [2012] 1 SLR 676. There are multiple reasons for the duty – that justice be seen to be done rationally, that decisions could be properly assessed for correctness on appeal, that the judge be forced to consider the factual and legal bases for his decision, that the parties leave the court knowing exactly why they have won or lost, and for the better information of the legal fraternity, among others.

120 Supreme Court of Singapore, *Judicial Code of Conduct for the Judges and Judicial Commissioners of the Supreme Court of Singapore* (20 February 2019) at pp 7–8.

the fairness that society both abides by and expects, to ensure that justice is both done and seen to be done.[121]

IV. AI and the judicial function

46 Having established the mode of operation of contemporary AI, as well as the scope of the judicial power, functions and duties, we are in a position to evaluate both the capability and the propriety of AI being used in judicial decision making.

A. Capability

47 As far as capability is concerned, one thing is eminently clear: as long as AI remains within the scope of weak AI, and as long as judicial decision-making retains its current qualitative approach, it is difficult to see how AI can approximate the entirety of judicial decision-making. The difference between the reasoning paths of AI (whether xAI or otherwise) and judicial decision-making is simply too great to be reconciled by a matter of mere advancement in present technology; a paradigm shift, whether in AI techniques or in legal reasoning, will have to take place before AI can do so.

48 Breaking down the various AI techniques available in the modern context, none of their ideal use cases model or replicate the conditions under which judicial decision-making is generally conducted. The simplest forms of AI, regressions, rely on statistical patterns observable in the data and deal badly with outliers, or data points so relatively outsized to the rest of the dataset that they skew its statistical distributions.[122] Slightly more sophisticated forms of AI, decision trees, are in effect a series of if-else conditionals writ large, which may result in overfitting of the model compared to the training data.[123] Modern AI techniques, using machine learning and artificial neural networks, are also vulnerable to outliers and corner case weaknesses; the dangers of such outliers are further compounded by the increased complexity of such techniques leading to difficulties in detecting and compensating

121 *R v Sussex Justices, ex parte McCarthy* [1924] 1 KB 256.

122 See, *eg*, Andreas Mielke, "Regression and Outliers" *Deloitte Digital* (11 February 2019) <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/technology/Deloitte-Trufa-ScienceInside4-Regression-Outliers.pdf> (accessed 23 July 2020).

123 George Seif, "A Guide to Decision Trees for Machine Learning and Data Science" *Towards Data Science* (30 November 2018) <https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956> (accessed 23 July 2020).

for these outliers.[124] All these techniques are necessarily bound by the data they are trained with, or solely endogenous observations, but this stands in stark contrast with judicial decision-making, which requires not just an understanding of specific laws and facts, but also a greater appreciation of the exogenous context in which laws are made and cases are decided. In so far as judges rely on previous cases, it is as an aid to tell them what the law is, and not exclusively to determine the range of possible factual inputs for the present case before them. A great deal of discretion by judges is still necessary to determine how to characterise the present case's inputs.

49 One further factor which restricts the use of AI in judicial decision-making is the present incapability of AI to comprehend the meaning of human language.[125] Current NLP techniques are restricted to reproducing a facsimile of human language by means of statistical analysis of large amounts of texts.[126] The lack of understanding is evidenced by the case of Tay, Microsoft's experimental Twitter chatbot, which within a matter of days if not hours had been corrupted by trolls on Twitter to spew racist hate speech, forcing Microsoft to take Tay down.[127] Tay's case is instructive in highlighting several pitfalls of AI in general: the principle of "garbage in, garbage out" where bad data corrupts outcomes, the need for constant oversight over an AI's processes to prevent corruption, and the fact that AI lacks any sort of moral compass or guide.

50 Lastly, any AI, subsisting solely on the data that is fed to it, is wholly incapable of taking cognisance of any material or information that is not contained within that data. AI systems therefore have to be explicitly programmed to be thusly cognisant of exogenous considerations on an *ad hoc* basis. While not every case requires cognisance of external information (*eg*, taking judicial notice,[128] or recognising policy imperatives such as the

124 See, *eg*, Takafumi Kanamori, Shuhei Fujiwara & Akiko Takeda, "Robustness of Learning Algorithms Using Hinge Loss with Outlier Indicators" (2017) 94 *Neural Networks* 173 at 173.

125 See, *eg*, Karen Hao, "AI Still Doesn't Have the Common Sense to Understand Human Language" *MIT Technology Review* (31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> (accessed 23 July 2020).

126 See, *eg*, Karen Hao, "AI Still Doesn't Have the Common Sense to Understand Human Language" *MIT Technology Review* (31 January 2020) <https://www.technologyreview.com/2020/01/31/304844/ai-common-sense-reads-human-language-ai2/> (accessed 23 July 2020).

127 See James Vincent, "Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day" *The Verge* (24 March 2016) <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist> (accessed 20 April 2020).

128 Jeffrey Pinsler, *Evidence and the Litigation Process* (Singapore: LexisNexis, 6th Ed, 2017) at para 11.002.

need for business certainty),[129] judges do not judge in a vacuum, and it is difficult to determine in advance whether or not exogenous information is needed to fully make sense of any particular piece of evidence or chain of legal reasoning. In addition, this implies that when laws are updated or changed, results output by an AI trained on older cases automatically becomes at least suspect if not irrelevant.[130] Such changes can include the creation of new laws, whether by Parliament[131] or by judges,[132] changes in sentencing frameworks put forth by judges,[133] or changes in the method of division of matrimonial assets.[134]

51 For the above reasons, we can safely conclude that the preconditions for AI to be used in a substantial manner in judicial decision-making do not currently exist. That being said, if developments in AI result in some form of true NLP – where an AI can actually interpret and replicate meaning in the same sense as humans do – then AI may well be capable of undertaking judicial decision-making.[135]

(1) Roles for AI in a smaller capacity

52 If AI cannot undertake the whole of the judicial function, can it still assist with aspects of judicial decision-making? Bearing in mind the limitations of AI as outlined above, AI could possibly assist in more standardised fields of law where there is sufficient data, where a premium is placed on consistency both with prior trends and future trends, and where precedent already exists for the use of mathematical reasoning, increasing the chances that the use of AI in such fields will be more readily accepted.

(a) Sentencing

53 One obvious example is the use of AI in sentencing scenarios. Here, as Menon CJ has indicated,[136] a number of crimes have simple and direct correlations between the gravity of the proscribed conduct and

129 See, *eg*, *Sarika Connoisseur Café Pte Ltd v Ferrero SpA* [2013] 1 SLR 531 at [61].

130 For a general statement of this principle relating to validation, see, *eg*, Melissa Hamilton, "Judicial Gatekeeping on Scientific Validity with Risk Assessment Tools" (2020) 38(3) *Behavioural Sciences & the Law* 226 at 227.

131 *Eg*, Misuse of Drugs Act (Cap 185, 2008 Rev Ed) ss 11B–11E.

132 *Eg*, *Malcomson Nicholas Hugh Betram v Mehta Naresh Kumar* [2001] 3 SLR(R) 379, where the common law tort of harassment was (temporarily) created.

133 *Eg*, *Ng Kean Meng Terence v Public Prosecutor* [2017] 2 SLR 449.

134 *Eg*, *ANJ v ANK* [2015] 4 SLR 1043.

135 The logical implication here is that humans have succeeded in creating strong AI, which has implications beyond the scope of the present discussion assuming weak AI.

136 *Ng Kean Meng Terence v Public Prosecutor* [2017] 2 SLR 449 at [30].

the length of the sentence handed down. In fact, the legal tech sector has already made advances in using AI to advise in sentencing scenarios, such as the work done locally by Lex Quanta in formulating outcome predictors for sentencing.[137]

54 Although it is more conceivable that AI can assist in sentencing as compared to general judicial decision-making, legal doctrinal issues need to be resolved before any questions of technical implementation. Chief amongst them is resolving the broad tension between retaining judicial discretion in sentencing and using mathematical formulae in sentencing. Given the mathematical nature of AI algorithms, using AI in sentencing will tend towards using mathematical formulae, whereas the present balance is more firmly in favour of judicial discretion, as evidenced by Menon CJ's disapproval in *Mohd Akebal s/o Ghulam Jilani v Public Prosecutor* of over-precision in sentencing.

55 It is not immediately clear how AI algorithms may replicate judicial discretion without incorporating a degree of randomness or arbitrariness. There is room for a legitimate multiplicity of judicial opinion as to what an appropriate sentence is in any particular case, given that a sentence must be manifestly excessive or inadequate before an appellate court will intervene, and not simply excessive or inadequate.[138] But this multiplicity of opinion is premised upon the same facts and cases cited to the courts. To achieve the same effect with AI, one would either have to use different training sets of data to achieve different predictive formulae, which would be equivalent to knowingly citing only some relevant case law but not all of it, or to add a random factor into the equation, which would be anathema to the concept of fair and transparent judicial decision-making. The solution may have to be to accept a greater degree of formulaic sentencing, which suggests that legal stakeholders have to be prepared to accept changes in the way that sentences are determined.

56 If AI is indeed brought in to assist in some aspect of judicial decision-making, the question is which model or models of xAI are most applicable and to what end. It may also well be that multiple types of xAI are required according to the needs of the various stakeholders involved. For instance, in the sentencing scenarios, legal practitioners could probably be satisfied with subject-centric explanations in order to know better how strong or weak their client's case is compared to historical examples. However, the Judiciary and court administrators deploying the xAI ought to have access also to model-centric explanations to ensure

137 Lex Quanta website <http://www.lexquanta.com/>.

138 See, *eg*, *Public Prosecutor v Mohammed Liton Mohammed Syeed Mallik* [2008] 1 SLR(R) 601 at [81]–[84].

that the AI has been appropriately deployed. While the source code may not be comprehensible to people without the requisite training, it should in principle also be made available to the Judiciary to ensure that they, with the necessary technical assistance, can satisfy themselves as to the propriety of the decisions made when challenges arise.

57 However, transparency is not the same as accuracy. An AI can be impeccably transparent yet hopelessly wrong. Naturally, other than ensuring that some form of xAI is chosen, relevant precautions and safeguards must be taken to ensure the accuracy of the AI to avoid the perversion of justice, and to ensure that the quality of decisions generated by AI is comparable to, if not better than, those rendered by human judges. The failure to do so is exemplified by the case of *Wisconsin v Loomis*,[139] which addressed the use (or misuse) of Correctional Offender Management Profiling for Alternative Sanctions ("COMPAS") in deciding whether or not to grant bail. It was discovered that COMPAS assessed black defendants to be more likely than white defendants to reoffend and therefore would assess them negatively. Unfortunately, the company behind COMPAS refused to divulge the details of its algorithm on the basis that these were proprietary.[140]

58 Without being able to examine the algorithm that COMPAS used, it is impossible to determine whether race was indeed explicitly considered a factor. For example, given the American historical context, race is likely to be correlated to socio-economic status, which itself is correlated to crime indicators such as type of crime and likelihood of reoffending. That way, likelihood of reoffending and race are but two manifestations of the same underlying symptom. But if the algorithm had explicitly used race as an indicator of likelihood of reoffending, such a link would likely run afoul of the judicial duty to be free from bias, as well as any clauses pertaining to constitutional equality.

(b) Judicial audits and outcome predictors

59 The example of COMPAS also suggests another role for AI in a capacity closely linked to judicial decision-making. In *Wisconsin v Loomis*, the defendant merely challenged COMPAS as a violation of his constitutional rights due to its inscrutability, and it was only in the course of examining the results produced by the AI in deciding that case that it was discovered that COMPAS was inclined to assess black defendants more harshly than white defendants.[141]

139 881 NW 2d 749 at [63] (2016).
140 *State of Wisconsin v Eric Loomis* 881 NW 2d 749 at [100] (2016).
141 *State of Wisconsin v Eric Loomis* 881 NW 2d 749 at [62] (2016).

60 Given that empirical legal analysis and AI share a common foundation of data science, the effort put into preparing legal big data for use by AI can also be tapped on for empirical legal analysis of existing cases, whether by the Judiciary itself as part of efforts at internal review, or by external parties and consultants like academics as a matter of research into legal decision-making.

61 In fact, prior to the actual deployment of AI in any judicial decision-making capacity, it is probably wise to conduct an empirical legal analysis of some sort as a precursor, in order to ensure the quality and consistency of the underlying decisions, and to bring any potential discrepancies to the awareness of the Bench and possibly provide time for the correction of any such discrepancies by the setting of new benchmarks.[142] Such analyses would also further illuminate the judicial decision-making process for the erudition of lawyers and, through them, the public. It bears noting that the use of such metrics is one way of making AI explainable by sketching out the rough parameters of the data that the AI operates within, although it will not be able to illuminate decisions in a particular case.

62 If used generally in the legal sector, this is one use of AI which has been explored for future use in the courts in the form of "outcome predictors".[143] In criminal cases, this is akin to the current practice of criminal case resolution conferences ("CCRs"), where a senior judge, based on his or her experience, would give an informal indication of the sentence that was likely to be appropriate in a quasi-mediation between the Prosecution and the Defence. Deploying AI in this scenario fulfils a dual purpose – it enables the intent behind CCRs to be more widely "available" to defence counsel for an indication that they can take back to their client,[144] and as an information/sentencing aggregator, it evens out the inequality of resources that defence counsel have been historically alleged to have suffered from.

63 There are of course questions about whether the use of outcome predictors would lead to "gaming" of the system. For instance, in the sentencing context, accused persons would potentially be able to calculate

142 See, *eg*, *Ng Kean Meng Terence v Public Prosecutor* [2017] 2 SLR 449 at [12]–[22], pointing out that the prior framework for sentencing in rape cases in *Public Prosecutor v NF* [2006] 4 SLR(R) 849 resulted in clustering of sentences and overall conceptual incoherence, necessitating a revision of the sentencing benchmark.

143 See *One Judiciary Annual Report 2018* at p 6.

144 This might assist counsel whose clients insist on some kind of official confirmation or indication before taking a certain course – this was anecdotally observed by the author during a criminal case resolution conference conducted by Principal District Judge Bala Reddy in August 2017, as part of his undergraduate coursework.

whether the additional jail time is worth the "risk" of claiming trial. While such tactical considerations may appear repugnant to the idea of justice and should in principle be discouraged, it is unlikely that these would result in perversions of the existing justice system. First, criminal lawyers are surely aware that accused persons receive a sentencing discount if they take a certain course, and should already be advising their clients accordingly. An outcome predictor merely provides this information in a more precise form, but has the added advantage of having the potential to provide this information to accused persons without the benefit of representation and hence the legal experience to know about a sentencing discount. Second, remorse is but one of the reasons to grant a sentencing discount, and there are other reasons to grant (or deny) sentencing discounts, such as if an accused chooses to plead guilty and potentially save court resources or spare the necessity of calling witnesses especially in sensitive cases like sexual assault cases.[145]

(2) Safeguards

64 If AI were indeed thus employed, measures must be taken to ensure that its decision-making process is fair and transparent. Such measures can take guidelines like the Personal Data Protection Commission's *Model AI Governance Framework* as a starting point. However, notwithstanding that the *Model AI Governance Framework* claims to be sector- and technology-agnostic, it must be remembered that the *Model AI Governance Framework* was released in the context of the adoption of AI solutions in a generic public context. Given the more consequential nature of judicial decision-making, it seems prudent to suggest that any safeguards implemented should be of a higher degree of rigour compared to the *Model AI Governance Framework* – if not in the number and nature of safeguards, then at least in their execution. Some examples of the degree of care with particular reference to the legal and judicial context follow.

65 First, ensuring the quality and holism of data fed into any legal AI application is of paramount importance. The scale of big data and the opacity of current AI techniques means that in the event of erroneous or skewed results returned by the AI, it is extremely difficult to figure out if the problem is caused by the data, the algorithms, or both. Having clean data for which the parameters are understood makes troubleshooting any AI application relatively much easier, and, as a matter of logic, increases the reliability of the results if the inputs can be shown to be statistically sound. In addition, the wider the ranges and combinations of fact patterns and decisions fed to the AI, the more robust the decisions rendered by the

145 See *Ng Kean Meng Terence v Public Prosecutor* [2017] 2 SLR 449 at [90].

AI will be. This implies that in the interests of maximising the accuracy of any AI deployed, access to cases and court records beyond the released grounds of decision available on LawNet is necessary. In the sentencing context, the State Courts were stated by the *One Judiciary Annual Report* to have dealt with over 300,000 criminal cases in 2018 alone; this is orders of magnitude above the 1,384 grounds of decision available for all types of cases on LawNet for 2018.[146] Furthermore, access to court filings is also required, as the AI must be trained to judge how to select suitable inputs for arriving at a correct output – this is not possible if an AI is trained only on grounds of decision for which judges have already filtered out irrelevant considerations from the filings.

66 Second, there must be awareness that there may be decisions made on fact patterns that fall outside the parameters of the AI's training data, and these must be caught and at least seriously examined if not wholly rejected, as they are at best unsafe and at worst entirely spurious. Such decisions are extrapolations based on mathematical trends that can be extracted from the training data, but which do not take into account logical or legal limits on these trends. For example, while it generally stands to reason that the punishment for crimes, such as the fine or jail term as the case may be, will increase proportionately with the severity of a crime, this is subject to the maximum punishment imposable by law. In cases where there are few to no instances of the maximum punishment being imposed, the AI may return a sentence based on a mathematical function derived from existing sentencing factors but which exceeds the statutory limit. While such instances are quite easily caught by simple, hand-coded checks against the relevant statute, it is precisely such cases that invite judicial consideration of whether the case indeed deserves the highest penalty possible, based on the existing rule of consideration of most serious instances of conduct.[147] It is unlikely that AI will be able to return a considered opinion based on mathematical formulae alone; indeed such an approach would be anathema to the courts which emphasise that judicial decision-making cannot be restricted thus.[148]

67 Third, significant human effort is required to pre-process and preliminarily evaluate the data that will form the basis of any AI engaged in judicial decision-making. This follows from the need to have clean data with well-understood parameters, as well as the nascence of NLP and its unreliability in consistently parsing meaning as we understand it.

---

146 The figure of 1,384 for LawNet cases in 2018 includes all decisions and judgments from the High Court, Court of Appeal and other tribunals. For State Courts cases, see *One Judiciary Annual Report 2018* at p 76.

147 See, *eg*, *Public Prosecutor v Govindasamy s/o Nallaiah* [2016] 3 SLR 374 at [96]–[97].

148 See, *eg*, *Dinesh Singh Bhatia s/o Amarjeet Singh v Public Prosecutor* [2005] 3 SLR(R) 1 at [21]–[24].

The human effort involved must be a cross-domain one – lawyers, data scientists, as well as lawyers trained as data scientists must be involved to ensure the quality of the data and to catch and process any biases that appear from the data.[149] It may well be that NLP techniques improve in time to the degree where its use may be considered in speeding up the parsing of legal texts, and it is possible that some form of legal texts, for instance statutes, obey sufficient conventions in their drafting that NLP techniques have less trouble in parsing them, but it would be folly to blindly trust that an AI application would parse human-origin human-purpose texts in the same way that we do.

68 Last, it also bears noting that withholding details about an AI algorithm used in judicial decision-making (like the company behind COMPAS) runs counter to the spirit of transparency and explainability otherwise expected. It is impossible to say whether the Judiciary will develop their own public solution or partner with a private entity to develop AI sentencing assistants at this point, but in both cases, some provision will need to be made for ensuring that the AI algorithm is transparent to at least the Judiciary and its experts, to ensure the continued accuracy of the algorithm.[150]

69 It is worth noting that some of these suggested safeguards have been elucidated in the Council of Europe's European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems,[151] in the form of the "principle of quality and security" and "principle of transparency, impartiality and fairness".[152] While some of the other principles in the aforementioned Charter are more applicable to a European rather than a Singapore legal context, the principles relating to the quality of data handling nonetheless apply equally across jurisdictions.

B. Propriety

69 While the above discussion suggests that we need not worry about AI replacing judges anytime soon, given the possibility (however improbable) that AI may in future be improved to the point where it

149 See Richard Susskind & Daniel Susskind, *The Future of the Professions* (Oxford: Oxford University Press, 2015) at pp 115–116.

150 Singapore Academy of Law Law Reform Committee, *Applying Ethical Principles for Artificial Intelligence in Regulatory Reform* (July 2020) at para 2.41.

151 Adopted at the 31st plenary meeting of the European Commission for the Efficiency of Justice (3–4 December 2018).

152 Council of Europe, European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (adopted at the 31st plenary meeting of the European Commission for the Efficiency of Justice (3–4 December 2018) at pp 10–11.

can undertake commonsensical reasoning well enough to approximate human, and therefore judicial decision-making, we must start to think about the propriety of AI undertaking judicial decision-making. Just because an AI can undertake judicial decision-making does not mean that it should. Is the nature of law, and of the judicial power that is vested in the courts and the Judiciary, one that can be practised by an AI? Does the fact that law began as a human construct, as a collation of rules that societies agree to live by, mean that it is only humans that can ever practice the law and judge others according to it? Does the maxim that justice must not only be done, but be seen to be done, mean that only a human judge can ever be seen to pass judgment? Such questions bring us away from strict legality into the realms of legal theory, human psychology, and computer ethics. While possibly new to law, such questions are familiar ones to computer scientists and ethicists plumbing the depths of what it means to be humanly intelligent and whether a computer can ever truly be described as intelligent in the same way as a human can be.

70 In a way, the question of whether an AI may ever hand down judicial decisions is similar to the thought experiment of the Chinese Room.[153] Briefly, the Chinese Room thought experiment involves a person locked in a room, with no communication with the outside world, tasked to translate English messages passed under the door into Chinese by following a provided algorithm. When the Chinese messages go out of the door, the people outside presume that there is someone within who understands Chinese, yet all the person inside has done is to unthinkingly follow rules. Can this person inside the room be said to understand Chinese? To delve into a critical analysis of the Chinese Room thought experiment is beyond the scope of this article, but the most common reply to the above postulate, known as the Systems Reply, is that while that person may not understand Chinese, he is instead part of a greater system that does understand Chinese and which has generated the rules for him to understand.

71 The legal analogue to the Chinese Room is to imagine two such locked rooms – in one sits a judge, and in another sits a computer. Identical cases are fed into each room, and identical decisions come back out. Contrary to the base Chinese Room, it is clear that both the judge and the computer are part of a greater legal system that has informed their decision-making. Modern legal systems have come a long way from the days where justice was arbitrarily the length of the chancellor's foot, and the legal doctrine of *stare decisis* controls and shapes common

153 See generally Stanford Encyclopedia of Philosophy, "The Chinese Room Argument" (revised 20 February 2020) <https://plato.stanford.edu/entries/chinese-room/#ChinRoomArgu> (accessed 20 April 2020).

law reasoning today. Does the provenance of the decision – whether it was made by a judge or a computer – affect our determination of the legitimacy of the decision, if both decisions are identical and made within the confines and tenets of the same system? Is the legitimacy of the decision premised only on following the rules of the system of justice that society expects, or the fact that it has been a human following that system? For routine cases, is there any principled, non-reactionary argument, from whichever sector or stakeholder in the legal system, against the use of intelligent agents to hand down decisions that a human would have made anyway?

72 It is difficult to see that there is any avenue of attack against the substantive correctness of a decision handed down by a machine if that decision was one that a human judge could indeed have handed down. The question is then whether any decision made by a sufficiently advanced AI can be attacked on the basis of procedural impropriety just because it was made by an AI (excluding any actual impropriety such as of the type feared in the use of COMPAS). Assuming that the AI's decision is sufficiently explainable and meets the requirements of judicial duties, present or future, such an attack should not be possible.

73 It may well be that despite our best efforts, AI will never be able to undertake the bulk of judicial decision-making, in a sufficiently explainable manner, to a degree that necessitates our needing to explicitly answer these questions of propriety. If human judges remain the touchstone for legal decisions, perhaps these questions are best dealt with as a matter of bureaucratic judicial operating procedure rather than as a larger question of legal ethics. That being said, if AI does assist in judicial decision-making, and it becomes known that AI is used in judicial decision-making, care must be taken to ensure that such use does not come to shake public confidence in judicial decision-making.[154] In this respect, it is not just legal stakeholders who must satisfy themselves as to the capability and propriety of the use of AI. Legal stakeholders must be able to engage the public and satisfy them as to the capability and propriety of such use, since it is the public that falls to be judged, in some part, by AI, and it would be desirable to avoid "regrettable and avoidable misconceptions" which overshadow broader points such as "the nature of the judicial mission and task".[155] But before legal stakeholders can do so,

154 See, *eg*, the robo-debt, Correctional Offender Management Profiling for Alternative Sanctions and social credit case studies in Monika Zalnieriute, Lyria Bennett Moses & George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82(3) Mod L Rev 425 at 436–440.

155 *Public Prosecutor v Siow Kai Yuan Terence* [2020] 4 SLR 1412 at [4] and [93]. The District Judge was unfortunately quoted out of context and oversimplified as stating that the accused got a relatively light sentence because he was well-educated; *(cont'd on the next page)*

they must begin the discussion amongst themselves as to what it means for an AI to assist in judicial decision-making, and that discussion has to begin now.

V. Conclusion

74 AI presently still has some way to go before it can be deployed reliably in any aspect of judicial decision-making. This article outlines the salient considerations that should be taken into account when supervising the preparation and deployment of AI in courts, requiring a thorough understanding of mathematics, data science and legal reasoning to be able to marry two otherwise disparate fields together. As the present state of the art in AI is not advanced enough to process cases in the same way as judges would, more work needs to be done in both doctrinal and technical fields in order to find ways to simulate legal thinking with AI, and also to advance the legal acceptability of mathematical reasoning. That said, mathematical reasoning is not by any means alien to courts, and there is potential for AI to be used in supporting roles in judicial decision-making, such as sentencing, judicial audits and outcome predictors. But there is still some way to go, and several prerequisites and safeguards to meet, before legal reasoning based purely on mathematical precepts can be safely accepted in their own right. On the one hand, such prerequisites imply that it is persons of great learning and eminence in these fields who must lay the groundwork on principled and defensible grounds, but on the other hand, the need for widespread confidence in our legal system means that such groundwork must also be easily explainable and accessible to laypersons who might otherwise have scant contact with these fields.

75 But this article is more than just a checklist of how and where to deploy AI in the courts. It is a call to action for legal stakeholders to begin asking themselves hard questions about the nature of the legal system that renders it favourable or otherwise to the use of AI, to have a shared discussion and conversation on the same, and finally to be confident and

Sundaresh Menon CJ clarified that the District Judge was quite rightly referring to his academic background as an indicator of his behaviour to assess if he had propensity for reform, although the sentence was subsequently increased because of fresh evidence and submissions before him that the District Judge had not had the benefit of.

competent in the use of AI, in whatever respect, in service to the law and the society which law itself serves.