Published on e-First 26 September 2019

TOWARDS A DIGITAL GOVERNMENT

Reflections on Automated Decision-making and the Principles of Administrative Justice*

Governments around the world are increasingly looking to utilise technology and automated systems in administrative decision-making. As Singapore undergoes its digital government transformation journey, automated decision-making may become an essential part of public service delivery in the near future. Such developments necessitate consideration of the interaction between automated decision-making and the principles of administrative justice. The first part of this article explains the types and benefits of automated decision-making systems. It then outlines the significant use cases by government agencies in Australia, Canada and the US, as well as the administrative law issues that these use cases illustrate. Drawing from these and other examples, the second part distils the key administrative law rules that may be infringed by the use of automated systems. The third part argues that such infringements can largely be avoided and suggests possible solutions. In the overall analysis, this article suggests there is much room for optimism that the use of automated systems for decision-making can remain consistent with the principles of administrative justice while enhancing public service delivery.

Makoto **HONG** Cheng**
LLB (Singapore Management University);
Advocate and Solicitor (Singapore);
Associate Fellow, Attorney-General's Chambers Academy.

* An earlier draft of this article was presented at a conference titled "Contemporary Issues in Public Law Litigation" at the Faculty of Law, National University of Singapore, on 9 January 2019. The authors would like to thank Yeong Zee Kin, Assistant Chief Executive, Info-communications Media Development Authority, for his invaluable comments and suggestions on the earlier draft. The authors are also grateful for the detailed and insightful comments from an anonymous peer reviewer. All errors and omissions remain the authors' own.

** The authors would like to acknowledge that the idea for this article originated from a presentation on the Singapore experience with automated decision-making delivered by Owi Beng Ki, Chief Legislative Counsel at the Attorney-General's Chambers, during the Australasian Parliamentary Counsel's Committee and CALC (Pacific) Drafting Conference held between 4 and 9 April 2018. The views expressed in this article are those of the authors and are not representative of the views of the Attorney-General's Chambers.

HUI Choon Kuen
LLM (New York University), MA (University of Oxford);
Advocate and Solicitor (Singapore);
Dean, Attorney-General's Chambers Academy.

I. Introduction

1 In May 2017, the world's reigning Go player, Ke Jie, was soundly defeated by "AlphaGo" – a program that used machine learning to master Go. While reminiscent of Gary Kasparov's defeat by IBM's chess playing supercomputer "Deep Blue" in 1997, experts hailed the latest event as a milestone in the development of artificial intelligence ("AI"), and in particular automated decision-making systems ("automated systems") built upon machine learning. Outside of such newsworthy events, developments in automated systems can be observed all around us. One example is how businesses are increasingly relying on customer support chatbots (which function autonomously in their interactions with customers) to address customer requests at all hours of the day. One success story is the chatbot developed by global fashion company H&M, which leverages on information provided by customers to help them navigate outfit possibilities and guide them to areas in H&M's online store that align with their desired purchases.[1] In recruitment, automated systems like "Mya" that can directly engage with candidates via text, ask and answer questions and rank candidates using weighted factors are being used to give recruiters and hiring managers more time to focus on other tasks like conducting interviews and closing offers.[2] In large part, such developments are driven by the need for companies to optimise resources, and expectations from customers for faster and more efficient service delivery.

2 These driving forces apply equally to the public sector, which is rapidly evolving to digitalise by using data, connectivity and computing to transform the way public services are delivered, and how public officers do their work. At the same time, the Government has recognised that digitalisation is a means to enrich the engagement between the Government and citizens, rather than an end in itself.[3] In

1 Daniel Faggella, "7 Chatbot Use Cases That Actually Work" *Emerj Artificial Intelligence Research* (14 December 2018).
2 Tonya Riley, "Get Ready, This Year Your Next Job Interview May be with an AI Robot" *CNBC* (13 March 2018).
3 Smart Nation and Digital Government Office, "Launch of Digital Government Blueprint: "A Singapore Government that is Digital to the Core, and Serves with Heart", press release (5 June 2018) at para 6 <https://www.smartnation.sg/whats-new/press-releases/launch-of-digital-government-blueprint--a-singapore-government-that-is-digital-to-the-core--and-serves-with-heart> (accessed 18 December 2018).

the same spirit, it is necessary to examine and address the issues that digitalisation may bring, including the interaction between automated systems and the principles of administrative justice.

3 This article seeks to contribute to this effort by initiating the conversation. The first part[4] explains the types and benefits of automated systems. It then outlines the significant use cases by government agencies in Australia, Canada and the US, as well as the administrative law issues that these use cases illustrate. Drawing from these and other examples, the second part distils the key administrative law rules that may be infringed by the use of automated systems, which are:

(a) illegality in the form of fettering of discretion, failing to consider relevant matters and considering irrelevant matters;

(b) irrationality; and

(c) improper delegation or devolution of powers, which is a subset of illegality but discussed separately as it involves a defect in the conferment of powers on a decision-maker rather than a defect in the decision *per se*.

4 The second part[5] also discusses a more fundamental question, which is whether a fully automated decision can be considered an administrative decision at all.

5 In the third part,[6] it is argued that potential infringements of the administrative law rules identified may largely be avoided by identifying the types of administrative decisions to be automated and designing automated systems with appropriate safeguards in place. It is suggested that illegality can be avoided by assessing the level of impact or harm that a decision will have on the rights or interests of the affected individual or entity, and instituting different levels of human intervention and monitoring as safeguards. Irrationality can be avoided by providing the appropriate level of disclosure that takes into account the state of technology, and balances administrative expediency with the need for rationality. The improper delegation of powers can be avoided by allowing a human decision-maker (assisted by an automated system) to retain control over a final decision, through legislative amendments to enable specific powers to be delegated to an automated system, or if devolution of a power to an automated system fulfils the two requirements of the *Carltona* principle.[7] However, whether it is

4 See paras 6–17 below.
5 See paras 18–39 below.
6 See paras 40–67 below.
7 See *Carltona Ltd v Commissioners of Works* [1943] 2 All ER 560.

appropriate to apply the concepts of delegation and devolution to automated systems at all remains a fundamental and open question that may remain unresolved until more sophisticated techniques are developed.

II. Overview of automated systems and use cases in Australia and Canada

A. Types and benefits of automated systems

6 As a starting point, an automated system refers to any information technology designed to produce measurements or assessments of a case about a particular individual meant to directly aid a human in his or her decision-making, or designed to make an administrative decision in lieu of a human decision-maker. Such systems process information in the form of input data using an "algorithm" to generate an output of some kind. An algorithm is a set of instructions, composed in programmable steps, designed for the purpose of organising and acting on a body of data to achieve a desired outcome.

7 Automated systems may be further classified into rule-based systems, and systems developed using machine learning ("machine learning system").[8] Rule-based systems are grounded in logic and rule-based programs that apply rigid criteria to factual scenarios, responding to input information entered by a user in accordance with predetermined outcomes. In comparison, machine learning is a method by which algorithms can be trained to recognise patterns within information, and the ways in which data interrelates. To develop an automated system through machine learning, algorithms are "trained" using an existing *corpus* of data (also called a training set), which enables an algorithm to classify and generalise beyond the examples in the training set. As the machine learns over time, it responds to feedback so that patterns learnt yield useful predictions or insights, and the machine's decisions are not based on rules that can be identified or explained. For this reason, a machine learning system may also be described as an "autonomous system" of decision-making.

8 A useful example to illustrate the distinction is the comparison between "Deep Blue" and "AlphaGo". On one hand, "Deep Blue" is a

---

8 For the purposes of this article, references to "machine learning" include references to neural networks. While machine learning traditionally uses statistical analysis to parse data and learn from that data to make decisions, neural networks are a class of machine learning model that analyses data with a view to mimicking human decision-making: see James Le, "A Gentle Introduction to Neural Networks for Machine Learning" *Codementor* (19 March 2018).

rule-based system designed to perform a complex tree search. At each point in the chess game, it would analyse the board and run an assessment on the possible moves it detected, and which would move it closer to a win. The key to the success of "Deep Blue" was its computing power, which enabled it to examine 200 million moves per second in the three minutes allocated for a single move in a chess game.[9] On the other hand, "AlphaGo" was designed to analyse millions of moves made by human Go experts, play games against itself and reinforce what it learned.[10] It was through this process that "AlphaGo" was able to master Go, which unlike chess has an almost infinite number of board arrangements and possible moves. This would not have been possible using the rule-based system that "Deep Blue" was based on.

9 Both rule-based and machine learning systems have the potential to bring significant benefits. For example, automated systems may be able to process eligibility decisions faster than, and as well as, a human in many circumstances by using appropriate program-related input data and a model to test inputs against rules.[11] This can enhance service delivery and manpower deployment by enabling applications to be processed at all hours, and more quickly than by humans.

10 From the perspective of administrative justice, automated systems have the capacity to reduce or eliminate the risk that a decision will be tainted by the motivations of the decision-maker (such as decisions made for an improper or ulterior purpose), or bad faith (such as decisions made arbitrarily, dishonestly or recklessly).[12] Automated systems also have the potential to make administrative decision-making more accurate, consistent and cost-effective, particularly where the decisions to be made are routine and voluminous. In addition, the ability of automated systems to make accurate predictions of complex phenomena can make decision-making more effective and efficient, and may help decision-makers avoid mistakes that cause a decision to be susceptible to judicial review. For example, an automated system may assist a decision-maker in the following ways:[13]

---

9 Gil Press, "The Brute Force of IBM Deep Blue and Google DeepMind" *Forbes* (7 February 2018).

10 Larry Greenemeier, "AI *versus* AI: Self-Taught AlphaGo Zero Vanquishes Its Predecessor" *Scientific American* (18 October 2017).

11 Treasury Board of Canada Secretariat, "Responsible Artificial Intelligence in the Government of Canada" Digital Disruption White Paper Series (10 April 2018) at p 17 <https://docs.google.com/document/d/1Sn-qBZUXEUG4dVk909eSg5qvfbpNlRhzIefWPtBwbxY/edit> (accessed 17 November 2018).

12 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 6.

13 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 3.

(a) guide a decision-maker through relevant facts, legislation and policy, closing off irrelevant paths along the way;

(b) provide useful commentary, including about relevant legislation, case law and policy, for the decision-maker at relevant points in the decision-making process; and

(c) help the decision-maker to identify correct questions to determine whether there are preconditions for the exercise of any powers, whether there exists evidence in respect of matters as to which the decision-maker must be satisfied, and particular issues for the decision-maker's consideration and evaluation.

B. Significant use cases in Australia, Canada and the US

(1) Australia

11 In Australia, rule-based systems are used extensively in assessing eligibility for social security payments. A person who receives welfare benefits updates information about his or her income and employment status fortnightly online. An automated system then processes the information to reassess the person's entitlement to social welfare payments.[14] In addition, the Australian Taxation Office ("ATO") has an online "e-tax" system to assist taxpayers in completing tax returns. With each click or selection from a telephone menu, the taxpayer is guided by the system through alternative pathways, and skips past options that the system determines are irrelevant. The process is completed with an assessment of the taxpayer's income tax or refund for that financial year, and the calculation of any penalties.[15]

12 Another example from Australia, which is illustrative of administrative law issues that may arise from the use of automated systems, is the Online Compliance Intervention debt recovery system ("OCI"). OCI is part of the Australian government's Better Management of the Welfare System initiative to recover $2.1bn of social security overpayments over four years. Introduced in July 2016, it automatically checked whether the income a person declares to Centrelink (a department within the Department of Human Services ("DHS") that disburses social security payments) matched ATO's records. Once a person has declared his or her income, or the deadline for entering information has passed, the automated system will calculate how much

14 Melissa Perry, "iDecide: The Legal Implications of Automated Decision-Making" [2014] FedJSchol 17 at p 4.

15 Melissa Perry, "iDecide: The Legal Implications of Automated Decision-Making" [2014] FedJSchol 17 at p 4.

money the person owes (if any) and send a debt recovery letter to the person if necessary.[16]

13 In 2017, OCI became the subject of a Senate Committee Inquiry and an investigation by the Ombudsman. As mentioned above, OCI matched the earnings recorded on a person's record with historical employer-reported income data from ATO. If a person did not respond to Centrelink, or if there were gaps in the information provided by the person, OCI automatically worked out a debt by "averaging" the total ATO recorded income across the period over which income was earned, and applied that average to each and every separate fortnightly calculation.[17] In other words, where data was incomplete, OCI used a person's assumed average income (and not the person's actual fortnightly income) to determine whether there was overpayment of social security payments. The algorithm assumed income is distributed evenly over 26 fortnights, with a presumption that work is either full time or part time, but with no provision for intermittent or casual work. This method of averaging income, and the fact that an application of the average may negatively affect the amount of debt owed, was not brought to the attention of many people. Further, DHS did not inform people that they could ask for an extension of time or be assisted by a human officer if they had problems. Because of this, the Commonwealth Ombudsman's Investigation Report (published in April 2017) was critical of the fairness and transparency of the decision-making process using OCI.[18]

14 The Ombudsman also criticised other aspects of OCI. It noted that one of the main efficiencies gained by OCI was that DHS no longer had to use its information gathering powers under ss 63, 192 and 195 of the Social Security (Administration) Act 1999. Instead, the onus was shifted to persons seeking social security benefits to provide the

16 National Social Security Rights Network, "Factsheet – Centrelink's Online Compliance Intervention" at p 1 <http://www.nssrn.org.au/wp/wp-content/uploads/2017/04/Centrelink-online-compliance-intervention.pdf> (accessed 15 November 2018).

17 Terry Carney Ao, "The New Digital Future for Welfare: Debts Without Legal Proofs or Moral Authority?" (2018) *University of New South Wales Law Journal Forum* 1 at 2.

18 Commonwealth Ombudsman, *Centrelink's Automated Debt Raising and Recovery System: A Report about the Department of Human Services' Online Compliance Intervention System for Debt Raising and Recovery* (Report No 2/2017, April 2017) at paras 3.14 and 3.15.

information.[19] This raised the issue of whether DHS had failed to exercise its information gathering powers.[20]

15 Further, the Ombudsman noted that there was a flaw in OCI's design that could have led to the fettering of a discretionary power to waive a 10% penalty for persons who refused or failed to provide information about their income, or had knowingly or recklessly provided incorrect information. According to the rules encoded in OCI, a case will be routed for manual application of the penalty if a person engages with OCI to indicate that there were personal circumstances that impacted the person's ability to declare his or her income. If not, the penalty will be automatically applied, and the person will receive a debt notification letter that provides the person with a further opportunity to provide reasonable excuse and have the fee removed. The Ombudsman questioned the effectiveness of this procedure in addressing the risk of fettering of discretion, opining that it could not have been if DHS did not effectively communicate the availability, meaning and importance of the reasonable excuse exception, and the ways of notifying the excuse to DHS.[21] In response to the Ombudsman's findings, DHS ceased the automatic charging of the 10% penalty fee, and now provides information on how a person can apply not to have the penalty imposed where he or she has a reasonable excuse.[22]

(2) Canada and the US

16 Canada has been developing automated systems to automate activities currently conducted by immigration officials, and to support the evaluation of immigrant and visitor applications, since 2014. As of

19 Commonwealth Ombudsman, *Centrelink's Automated Debt Raising and Recovery System: A Report about the Department of Human Services' Online Compliance Intervention System for Debt Raising and Recovery* (Report No 2/2017, April 2017) at para 2.5.

20 This may also account for why the Senate Community Affairs Committee emphasised during the Senate Committee Inquiry that the onus of proving a debt must remain with the Department of Human Services ("DHS"). The Committee also recommended that DHS resume full responsibility for calculating verifiable debts (including manual checking), based on actual fortnightly earnings and not an assumed average: see Terry Carney Ao, "The New Digital Future for Welfare: Debts without Legal Proofs or Moral Authority?" (2018) *University of New South Wales Law Journal Forum* 1 at p 4.

21 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 20–21; Commonwealth Ombudsman, *Centrelink's Automated Debt Raising and Recovery System: A Report about the Department of Human Services' Online Compliance Intervention System for Debt Raising and Recovery* (Report No 2/2017, April 2017) at paras 2.36–2.43.

22 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 22.

2018, Immigration, Refugees and Citizenship Canada is using an automated system to "triage" certain applications into two streams, with "simple" cases being processed by the system and "complex" cases being flagged for human review.[23] Another agency that uses automated systems is Employment and Social Development Canada, the body responsible for administering employment insurance and benefits. It uses a risk-scoring algorithm that is said to have "greatly improved the effectiveness of its overpayment investigations".[24] The Department of Justice Canada is also considering the use of automated predictive models to manage legal risks and allocate resources for legal service delivery.[25]

17 Such automated predictive models are already prevalent in the US in the form of risk assessment tools that are used to make recommendations for critical decisions like arrest, detention and supervision, based on a statistical prediction of the likelihood of a person committing an offence in the future.[26] They also illustrate the problem of algorithmic bias when automated systems are used to make decisions. An example is COMPAS, an algorithm used to guide sentencing by predicting the likelihood of a criminal reoffending. In 2016, COMPAS was reported to be racially biased, with a study published in January 2018 claiming that COMPAS is no better than a random group of people with little to no criminal justice experience in accurately predicting who is likely to reoffend.[27] Similar criticism was made of PredPol, a predictive policing software designed to predict when and where crime will take place, with the aim of mitigating human bias in policing. Used in several US states, PredPol was found by researchers to have sent officers to neighbourhoods with a high proportion of people from racial minorities, regardless of the true crime rate in those areas. Researchers also demonstrated that because the

---

23 Petra Molnar & Lex Gill, "Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System" (Faculty of Law, University of Toronto) (September 2018) at p 14 <https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf> (accessed 17 November 2018).

24 Shared Services Canada, Steering Committee on Big Data, "Diagnostic Report", released under the Access to Information Act to Lex Gill (Citizen Lab) (8 December 2014) at p 8 <https://drive.google.com/file/d/1HpVgzMdF7SPH319iNA2wkkQkDwDax7F7/view?usp=sharing> (accessed 14 November 2018).

25 Shared Services Canada, Steering Committee on Big Data, "Diagnostic Report" (8 December 2014), released under the Access to Information Act to Lex Gill (Citizen Lab) at p 8 <https://drive.google.com/file/d/1HpVgzMdF7SPH319iNA2wkkQkDwDax7F7/view?usp=sharing> (accessed 14 November 2018).

26 John Raphling, "Human Rights Watch Advises against Using Profile-based Risk Assessment in Bail Reform" *Human Rights Watch* (17 July 2017).

27 Julia Dressel & Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism" (2018) 4(1) *Science Advances* at p 3.

software learns from reports recorded by the police rather than actual crime rates, PredPol creates a "feedback loop" that can exacerbate racial bias.[28]

III. Key administrative law rules that may be infringed

A. Illegality

(1) Fettering of discretion

18 As mentioned above, rule-based systems apply rigid criteria to factual scenarios. This does not accord with the reality of many administrative decisions, which require evaluative judgment and the exercise of discretion. For example, a decision-maker may need to determine whether there is "good reason" to require a corporation to produce books relating to its affairs,[29] or whether making an order to prevent a person from entering a casino would be in the best interests of the person and his or her family members.[30] These are arguably questions that cannot be transcribed into rigid criteria or rules. If rule-based automated systems are used to make decisions involving such questions, the decision may be challenged on the basis that the automated system had failed to exercise discretion by applying predetermined outcomes that cannot be deviated from.[31] In other words, the automated system had fettered its discretion by not exercising discretion at all.

19 In Singapore, the administrative law rule against fettering of discretion ("no-fettering rule") requires a decision-maker to exercise genuine discretion and be prepared to hear out individual cases or deal with exceptional cases.[32] This means that even though a decision-maker may follow a policy or guideline that is within the scope of its conferred powers, the decision-maker cannot rigidly exclude the possibility of any

28 Matt Reynolds, "Biased Policing is Made Worse by Errors in Pre-Crime Algorithms" *New Scientist* (27 April 2018).

29 See s 8A(1) of the Companies Act (Cap 50, 2006 Rev Ed), which enables the Minister to give directions to a corporation to produce such books relating to its affairs as may be specified, if the Minister is satisfied that there is good reason to do so.

30 See s 162(1) of the Casino Control Act (Cap 33A, 2007 Rev Ed), which enables a Committee mentioned in s 159(2) of that Act to make a "family exclusion order" against a person if, among other things, the Committee is satisfied that doing so would be in the best interests of the person and his or her family members.

31 Melissa Perry, "iDecide: The Legal Implications of Automated Decision-Making" [2014] FedJSchol 17 at p 5.

32 *Lines International Holding (S) Pte Ltd v Singapore Tourist Promotion Board* [1997] 1 SLR(R) 52 at [78].

exception to that policy or guideline in deserving cases.[33] Persons affected by a decision must also be given the opportunity (whether via a hearing or any other method of making representations) to persuade the decision-maker to amend or deviate from the policy or guideline.[34] Otherwise, the decision may be challenged on the judicial review ground of illegality, which prohibits a decision-maker from incorrectly understanding the law regulating his decision-making power or failing to give effect to it.[35]

20 In the authors' view, compliance with the no-fettering rule would either require decisions to be made only by automated systems capable of exercising discretion, or, if rule-based automated systems are used, the involvement of human decision-makers at certain stages of the decision-making process. For the reasons set out below,[36] it is suggested that the latter method is preferable.

(2) Failing to consider relevant matters and considering irrelevant matters

21 The examples of COMPAS and Predpol discussed above[37] illustrate the issue of "algorithmic bias" that may occur when machine learning systems are used to make decisions. Despite the similarity in terminology, algorithmic bias is conceptually distinct from the administrative law rule against bias, which together with the hearing rule comprise the administrative law rules of natural justice.[38] As the administrative rule against bias prohibits a person from acting as a judge in his own cause in any matter where there is an actual or potential conflict of interest,[39] a machine learning system making a decision would need to have a distinct "interest" (separate from that of its human designer) before the rule may be infringed. It is doubtful that this would

33 Lord Woolf *et al*, *De Smith's Judicial Review* (Sweet & Maxwell, 8th Ed, 2018) at para 9-004.
34 *Lines International Holding (S) Pte Ltd v Singapore Tourist Promotion Board* [1997] 1 SLR(R) 52 at [78]; Lord Woolf *et al*, *De Smith's Judicial Review* (Sweet & Maxwell, 8th Ed, 2018) at para 9-004.
35 *Tan Seet Eng v Attorney-General* [2016] 1 SLR 779 at [79], citing *Council of Civil Service Unions v Minister for the Civil Service* [1985] AC 374 at 410–411. There is suggestion that a decision that infringes the no-fettering rule may also be challenged on the judicial review ground of procedural impropriety, as the affected person is precluded from influencing the use of that discretion or participating in the decision-making process: Lord Woolf *et al*, *De Smith's Judicial Review* (Sweet & Maxwell, 8th Ed, 2018) at para 9-003. However, there is no clear pronouncement in Singapore case law to this effect.
36 See paras 40–67 below.
37 See para 17 above.
38 *Yong Vui Kong v Attorney-General* [2011] 2 SLR 1189 at [88].
39 *Yong Vui Kong v Attorney-General* [2011] 2 SLR 1189 at [91].

occur unless the machine learning system possesses the same sentience as a human decision-maker. What algorithmic bias in fact does is to cause machine learning systems to give disproportionate weight to some factors and excessive weight to other factors, fail to consider relevant issues that a decision-maker is required to consider, or consider irrelevant issues, depending on how the system is designed to differentiate and sort data. A decision made by a machine learning system affected by algorithmic bias may be challenged on the judicial review ground of illegality, which prohibits a decision-maker from failing to take account of relevant considerations or taking account of irrelevant considerations when exercising a power ("rule on relevant considerations").[40]

22 Algorithmic bias may occur at the problem framing stage, data collection stage and the data preparation stage.[41] At the problem framing stage, failure to accurately define what the machine learning system is intended to compute may result in the system reflecting the designer's prejudices or pre-existing societal biases, or even embody values that the designers did not intend.[42] A hypothetical example is of a machine learning system to predict whether a person is a "fit and proper person" to hold a certain licence in a male-dominated industry. As the phrase "fit and proper" is open textured, the designer must decide, for example, whether the policy objective is to limit licences to persons who possess specific expertise, or persons who are of good character. This would then determine the type of data to be collected and used. If the policy objective is the former, the dataset may include the professional qualifications of all past applicants for a licence for the machine learning system to compare a new applicant against. If the policy objective is the latter, the dataset may include data that is indicative of character from a much larger group of people for such comparison to be made. In this example, using the former dataset when the policy objective is the latter may result in eligible female applicants being systematically filtered out even though this may not be what the designer intended.

23 Algorithmic bias may also occur at the data collection stage if the data collected is either unrepresentative of reality or reflects existing prejudices. An example is the recruiting tool used by Amazon, which was trained using historical hiring decisions by Amazon that favoured men, and learnt to show the same bias.[43] At the data preparation stage,

40 *Tan Seet Eng v Attorney-General* [2016] 1 SLR 779 at [80].
41 Karen Hao, "This Is How AI Bias Really Happens – And Why It's So Hard to Fix" *MIT Technology Review* (4 February 2019).
42 Daniel Cossins, "Discriminating Algorithms: 5 Times AI Showed Prejudice" *New Scientist* (12 April 2018).
43 Forbes Insights with Intel AI, "Overcoming AI's Challenge in Hiring: Avoid Human Bias" *Forbes* (29 November 2018).

where the designer must choose which attributes in the dataset the machine learning system will consider, bias may occur if the wrong attributes are chosen or if other attributes are ignored. In the licensing example above, selecting (or failing to omit) the continuous number of years an applicant has spent in the industry as an attribute to determine whether the applicant possesses specific expertise may result in bias against otherwise qualified and experienced applicants who left the industry for a period in between their careers to pursue other interests.

24 There is growing recognition in the private sector that algorithmic bias exists in many areas, and that the rapid uptake of AI for tasks like filtering candidates in hiring has moved too quickly with too little scrutiny. There are also views that algorithms cannot simply be trained on historical data without first examining the assumptions and biases embedded in a dataset,[44] and removing data in the dataset that can carry algorithmic bias.[45] Such concerns are equally salient in administrative decision-making, especially when it comes to what factors to give greater or lesser weight to, and what considerations are more or less relevant.

B. Irrationality

25 Rationality in the exercise of official powers, which seeks the accuracy of decisions and prohibits arbitrariness, is a key pillar of administrative law and an important part of the rule of law.[46] Irrationality as a ground of judicial review differs from illegality in that while illegality examines whether a decision-maker has exercised his discretion within the scope of his authority, irrationality looks at whether a decision, even though falling within a range of legally possible answers, is "so unreasonable that after considering the correct factors, no reasonable decision-maker could have come to it".[47]

26 Decisions made by machine learning systems may be challenged on grounds of irrationality if the steps leading to the decision are not readily accessible, which in turn makes it difficult or impossible to determine whether or not the decision was rational. Consider Lord Greene MR's famous example of a teacher who was dismissed solely on the basis that she had red hair to illustrate a decision that

44 Eric Rosenbaum, "Silicon Valley Is Stumped: Even AI Cannot Always Remove Bias from Hiring" *CNBC* (30 May 2018).
45 Forbes Insights with Intel AI, "Overcoming AI's Challenge in Hiring: Avoid Human Bias" *Forbes* (29 November 2018).
46 Lord Woolf *et al*, *De Smith's Judicial Review* (Sweet & Maxwell, 8th Ed, 2018) at para 1-025.
47 *Tan Seet Eng v Attorney-General* [2016] 1 SLR 779 at [80].

would be considered irrational.[48] It may be assumed that in the example, the teacher had challenged the decision and the decision-maker was asked to provide justification, which led to the revelation of the reason for the decision. If the decision had instead been made by a machine learning system whose reasoning process and considerations are unexplainable, the court will probably find (in the absence of conceivable reasons for the teacher's dismissal) that the decision was made without any legal basis and therefore both illegal and irrational.

27 The reason for the opaqueness of machine learning systems may be that the algorithm or some inputs are secret, the implementation is secret, or the decision-making process is not precisely described.[49] Unlike rule-based systems, which use deterministic rules and a combination of "if-then" statements that guide a computer to reach a conclusion or recommendation, machine learning systems assume that outputs can be described by a combination of input variables and other parameters. These outputs could be a simple binary classification such as whether one should bring an umbrella out, or something more complex such as the likelihood of a transaction being fraudulent. In both cases, the output is determined by a neural network. A network's reasoning is embedded in the behaviour of thousands of simulated neurons, arranged into layers. The neurons in the first layer each receive an input, and perform a calculation, before outputting a new signal. This is then fed to the neurons in the next layer, and so on, until an overall output is produced. Adding to the complexity is a process known as back-propagation, which tweaks the calculations of individual neurons in a way that lets the network learn to produce a desired output. The interplay of calculations using these mechanisms, which is crucial to complex decision-making, is also often opaque and inscrutable.[50] This is why an automated system developed through machine learning is also called a "black box" – one may observe inputs and outputs, but what really happens to the inputs occurs "in the dark". In completely automated systems, every input and most outputs are also encapsulated in the "black box", such that the information sources representing inputs are not observable.[51]

28 In the context of discretionary decision-making, concerns have been raised that while a machine learning system may be capable of

48 *Associated Provincial Picture Houses Ltd v Wednesbury Corp* [1948] 1 KB 223 at 229, citing *Short v Poole Corp* [1926] Ch 66 at 91.

49 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 2.

50 Will Knight, "The Dark Secret at the Heart of AI" *MIT Technology Review* (11 April 2017).

51 Dominique Hogan-Doran, "Computer Says 'No': Automation, Algorithms and Artificial Intelligence in Government Decision-making" (2017) 13 TJR 1 at 31.

exercising a discretionary function, the reasoning behind the machine's decision may be inscrutable. This is because the steps leading to that decision "would not readily be accessible" and "may be no more translatable into logic and reasoning than the mental processes of the human decision-maker where logic is exhausted and only discretion remains".[52] The problem caused by "black boxes" has led to recommendations by AI experts in the US that core public agencies, such as those responsible for criminal justice, healthcare, welfare and education and other "high stakes" domains, no longer use "black box" AI and algorithmic systems. This is because the use of such systems by public agencies raises serious due process concerns. The experts also recommended that, at a minimum, such systems should be available for public auditing, testing and review, and be subject to accountability standards.[53]

C. Improper delegation or devolution

29 It is a principle of administrative law that when Parliament vests power in an official or authority, there is a presumption that the power must be exercised by that named official or authority ("principal"), and no other person. While this presumption may be rebutted by contrary indications found in the language, scope or object of the statute,[54] it is usually strictly applied even where administrative inconvenience may result.[55] In order for a power to be validly exercised by a person other than the principal, the power must either be delegable or capable of devolution, and be properly delegated or devolved to that person. Otherwise, the exercise of that power by the person to whom the power was purportedly delegated or devolved may be challenged on the judicial review ground of illegality. There is no reason why this should not also apply where an automated system is utilised to make part or all of a decision, either as a delegate of the principal or a "person" to whom the principal's power is devolved ("devolvee").

30 Where the potential delegate or devolvee is a natural person, Parliament and the courts have developed both statutory and common law methods for statutory powers to be delegated or devolved to that

52 Robert French, "Rationality and Reason in Administrative Law – Would a Roll of the Dice Be Just as Good?" Australian Academy of Law Annual Lecture (29 November 2017) at p 4.

53 Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, "AI Now 2017 Report" at p 1 <https://ainowinstitute.org/AI_Now_2017_Report.pdf> (accessed 10 February 2019).

54 Lord Woolf *et al*, *De Smith's Judicial Review* (Sweet & Maxwell, 8th Ed, 2018) at para 5-160.

55 William Wade & Christopher Forsyth, *Administrative Law* (Oxford University Press, 11th Ed, 2014) at p 259.

person. These include delegation under s 36(1) of the Interpretation Act,[56] specific statutory provisions that expressly enable delegation, and the principle derived from the case of *Carltona Ltd v Commissioners of Works*.[57] As these statutory and common law methods were not designed with automated systems in mind, there is a risk that an exercise of statutory powers by automated systems as a delegate or a devolvee may be found to be *ultra vires*.

D. Whether fully automated decisions are decisions

31 Other than the administrative law issues discussed above, the decision by the Full Federal Court of Australia in *Pintarich v Deputy Commissioner of Taxation*[58] ("*Pintarich*") raises a more fundamental question of whether decisions made by automated systems are decisions at all. If they are not, then the issue of whether such a decision would pass muster under administrative law rules would be moot in the absence of any decision to begin with.

(1) Pintarich v Deputy Commissioner of Taxation

32 In *Pintarich*, a taxpayer received a computer-generated letter from ATO that ostensibly waived most of the general interest charge ("GIC") on a tax debt. Subsequently, ATO advised the taxpayer that he was in fact liable to pay the tax debt. The taxpayer applied to the Federal Court for judicial review but was unsuccessful. On appeal, the majority of the Full Federal Court found that no decision had been made as the automated letter was not accompanied by the requisite mental process of an authorised officer.

33 The majority of the Full Federal Court agreed that based on a natural reading of the computer-generated letter, the Deputy Commissioner of Taxation had agreed to accept a payment of a specified lump sum in full discharge of the taxpayer's primary tax and GIC liabilities. However, the majority did not think that this meant the Deputy Commissioner had made the decision. The majority was persuaded by the reasoning of Finn J in *Semuningus v Minister for Immigration and Multicultural Affairs*[59] ("*Semuningus*") that two elements must be satisfied before a decision is valid. First, there must be a "mental process" of reaching a conclusion. Second, there must be an objective manifestation of that conclusion. On the facts, the majority

56 Cap 1, 2002 Rev Ed.
57 [1943] 2 All ER 560.
58 [2017] FCAFC 79.
59 [1999] FCA 422 at [19], affirmed on appeal in *Semunigus v Minister for Immigration and Multicultural Affairs* (2000) 96 FCR 533 at [11], [55] and [101].

found that the decision was reached by an officer to whom the Deputy Commissioner's power was devolved by keying in certain information into a computer-based "template bulk issue letter". The majority found that as this produced a letter that did not reflect the officer's intentions in some respects, and the letter did not expressly deal with the application to remit GIC, neither the Deputy Commissioner nor the officer had reached a conclusion as to the application for remission of GIC.[60]

34 On the other hand, the dissenting judge in *Pintarich* (Kerr J) observed that the legal conception of what constitutes a decision cannot be static, and must take into account the fact that technology has altered how decisions are in fact made and that some or all aspects of decision-making can occur independently of human input.[61] To Kerr J, whether a decision has been made must be fact and context specific, requiring close assessment of whether the circumstances in which the conduct that is allegedly a decision was "within the normal practices of the agency and whether the manifestation of that conduct by an overt act would be understood by the world at large as being a decision".[62] Kerr J also considered that it would undermine fundamental principles of administrative law for a decision-maker to renounce a decision or purported decision by asserting that there was a distinction between the decision-maker's mental processes and the expression of those mental processes in an overt act (both of which are required by *Semuningus*).[63]

(2) B2C2 Ltd v Quoine Pte Ltd

35 In comparison to *Pintarich*, the Singapore International Commercial Court ("SICC") adopted a more pragmatic approach in the recent case of *B2C2 Ltd v Quoine Pte Ltd*[64] ("*B2C2*"). In *B2C2*, the plaintiff and defendant traded two types of cryptocurrency, Bitcoin and Ethereum, with each other using algorithmic trading software that executed trades with minimal human input. Human traders were almost never involved in the trading process, and human intervention was mostly limited to "tasks like monitoring, IT maintenance and adjusting the parameters of the Trading Software from time to time".[65] The trading software was based on rule-based systems as it was deterministic, had no mind of its own and only did what it had been programmed to do.[66]

60 *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 at [152]–[153].
61 *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 at [46].
62 *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 at [51]–[52].
63 *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 at [55].
64 [2019] 4 SLR 17.
65 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [13].
66 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [82] and [208].

36 The seven disputed trades involved the defendant's trading system erroneously crediting Bitcoin to the plaintiff in exchange for Ethereum at around 250 times the prevailing exchange rate. One of the defendant's arguments was that the trades should be either void or voidable on the basis of unilateral mistake at common law and in equity. However, it was not contended, and the court did not find, that a "decision" for each of the disputed trades had not been made because there was no "mental process" by a human decision-maker in the making of each trade. This appears to be both sensible and reflective of commercial realities because if such "mental process" were required for every trade, the speed and utility of algorithmic trading software are likely to be significantly diminished.

37 While the absence of human intervention in *B2C2* did not affect the validity of the disputed trades, the court had to decide *who* must have made a relevant mistake at common law or in equity in the context of algorithmic trading software, and how knowledge of that mistake could be ascertained.[67] On the first issue, the court held that the relevant mistake must be "a mistake by the person on whose behalf the computer placed the order in question as to the terms on which the computer was programmed to form a Trading contract in relation to that order".[68] On the second issue, the court held that in so far as rule-based systems are concerned, knowledge of a mistake should be ascertained with regard "to the state of the mind of the programmer of the software of that program at the time the relevant part of the program was written".[69] As the programmer did not have actual or constructive knowledge of the defendant's mistaken belief, the defence of unilateral mistake failed at both common law and in equity.[70]

38 In the authors' view, the pragmatic approach of Kerr J in *Pintarich* and the SICC in *B2C2*, which recognises that the legal conception of what constitutes a decision should evolve to reflect the reality of how decisions are made in the age of digitalisation, is clearly preferable to the rigid approach of the majority in *Pintarich* that suggests discretion can only be exercised by a human decision-maker. The method of attribution used in *B2C2* can be usefully applied in administrative law to determine whether decisions made by rule-based systems infringe any administrative law rules. Possible use cases may include the following:

(a) ascertaining whether a decision was made in bad faith or with bias by looking into the programmer's subjective state of

67 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [198].
68 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [205].
69 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [211].
70 *B2C2 Ltd v Quoine Pte Ltd* [2019] 4 SLR 17 at [231] and [236].

mind or objective when programming the system, using evidence such as the logic and decision-making pathways of the system; and

(b) ascertaining whether a rule-based system had considered irrelevant matters, or failed to consider relevant matters, using evidence such as the data used by the programmer and how the system was programmed to weigh different factors.

39 The limitation of this method of attribution is that it cannot be used for decisions by machine learning systems, which gradually take on minds of their own as they teach themselves with datasets provided by designers. In such cases, there would be no human to whom any state of mind may be attributed. Looking ahead, new legal rules that recognise this possibility, and techniques to ascertain the states of mind of machine learning systems when making decisions autonomously, may need to be developed.

IV. Possible solutions

A. Illegality: Limit types of decisions by automated systems

(1) Compliance with no-fettering rule

40 To comply with the no-fettering rule, a possible solution is to limit automated decision-making to decisions that only involve objective elements. This approach has been advocated by jurists who believe that if automated systems were used to exercise a discretion or make an evaluative judgment, there would be a constructive failure to exercise the discretion because the automated system would be applying predetermined outcomes that may be characterised as prejudgment or bias.[71]

41 Applying this solution entails that where a decision is a result of a process involving both objective and subjective elements, the automated system would be designed to only determine the objective elements of the decision. Take the hypothetical example of s 37(6) of the Environmental Public Health Act,[72] which enables the Director-General of Public Health to revoke or suspend the licence of a food establishment, private market or hawker if any of the following criteria are met:

71 Melissa Perry, "iDecide: The Legal Implications of Automated Decision-Making" [2014] FedJSchol 17 at p 5.
72 Cap 95, 2002 Rev Ed.

(a) the licensee is suffering from an infectious disease;

(b) the licensee knowingly employs any person who is suffering from or is suspected to be suffering from an infectious disease;

(c) the licensee, his assistant or employee refuses to comply with any requisition made by the Director-General under s 37(1), 37(2) or 37(3); and

(d) the licensee fails to ensure that his assistant or any person employed by him is immunised against any infectious diseases as required by the Director-General.

42 Among these criteria, (a), (c) and (d) may be objectively determined, while (b) is subjective in nature because it requires the Director-General to be satisfied of the licensee's knowledge. The solution would enable criteria (a), (c) and (d) to be determined by an automated system, while reserving criterion (b) for determination by a human decision-maker.

43 It is argued that while this solution may appear to be straightforward, determining whether or not a decision involves subjective elements requires more than just interpreting the legislative text. For example, the Director-General and his officers may have developed a checklist of factors using historical data to determine whether a licensee has the requisite knowledge to meet criterion (b). A junior officer may be tasked with evaluating each case using the checklist and identifying special factors on the facts that are not covered by the checklist. If there are no special factors, the junior officer may make a decision based on the checklist. If there are special factors, the junior officer would have to refer the decision to a senior officer for further evaluation.

44 In the authors' view, deploying a rule-based system to determine whether criterion (b) is met in the above circumstances is in principle no different from tasking a junior officer with making the initial decision. Doing so is therefore unobjectionable as long as the decision-making framework is designed with administrative law requirements in mind. In the first scenario where the automated system discovers no special factors and makes a decision using the checklist, the no-fettering rule can be complied with by administratively allowing licensees to write in and raise new factors for the Director-General's consideration. A reconsideration of the automated system's decision may then be made by a human decision-maker. In the second scenario where the automated system discovers special factors, the no-fettering rule can be complied with by requiring the automated system to refer the

decision to a human decision-maker, just as a junior officer would to a senior officer.

45 The discussion above illustrates two points. First, it is necessary (regardless of the language of a statute) to inquire into whether a rule-based system is exercising subjective judgment, or is merely assisting human decision-makers to process information and arrive at preliminary decisions based on predetermined guidelines that can be applied by considering whether objective elements are satisfied – in the same way that the human decision-maker would have done manually. Second, it is possible for decisions of rule-based systems to pass muster under the no-fettering rule if the decision-making framework is designed to involve human intervention (wherever necessary) at appropriate stages of the decision-making process, and if there is a mechanism for the escalation of exceptions for handling by human decision-makers.

(2) Compliance with rule on relevant considerations

46 It may be argued that limiting automated systems to only making decisions with objective elements will lead to compliance with the rule on relevant considerations. This is because algorithmic bias cannot arise where a decision does not contain subjective elements. In the above example, an argument against allowing machine learning systems to determine (b) would be that the automated system may produce and magnify biased outcomes unless the training datasets are first examined and adjusted for factors that are indicative of bias. For example, suppose historical data shows that the "knowledge" requirement in (b) is typically met in the case of hawkers who operate from stalls, and where both the hawker and the stall assistant speak Mandarin. In comparison, the requirement is seldom met in the case of proprietors of food establishments, and where both the proprietor and his or her employees speak English. Reliance on such data without accounting for possible bias against hawkers, and persons who speak Mandarin, may create a self-fulfilling feedback loop that makes it much more likely for Mandarin-speaking hawkers to be found to have had knowledge, *vis-à-vis* English-speaking proprietors of food establishments. This would not occur if machine learning systems are precluded from determining (b).

47 Despite the benefits of this solution, the authors do not think that it is the only solution to comply with the rule on relevant considerations. Further, the risk of algorithmic bias does not mean that it is necessary to preclude machine learning systems from making decisions with subjective elements altogether. An alternative solution is discussed below.

B. Illegality: Categorise decisions by risk level and implement appropriate safeguards

48 It is suggested that an alternative solution to comply with the rule on relevant considerations is to enable automated systems to also make decisions with subjective elements, but require the system designers (working in conjunction with human decision-makers vested with statutory powers) to classify decisions that the system will make by the level of impact, or degree of harm, that the decision will have on the rights or interests of the affected individual or entity. Depending on the level of impact or harm, different levels of safeguards that range from human intervention and monitoring to specifying factors considered in making the decision may then be instituted.

49 In Canada, a Treasury Board Directive on automated decision-making (which is in development) requires an "Algorithmic Impact Assessment" to be completed prior to the production of any automated system. This assesses the impact that a decision made by the automated system will have on (a) the rights of individuals or communities; (b) the health or well-being of individuals or communities; (c) the economic interests of individuals, entities or communities; and (d) the ongoing sustainability of an ecosystem. The potential impact is categorised into four levels:[73]

(i) Level I decisions will often lead to impacts that are reversible and brief.

(ii) Level II decisions will often lead to impacts that are likely reversible and short-term.

(iii) Level III decisions will often lead to impacts that can be difficult to reverse and are ongoing.

(iv) Level IV decisions will often lead to impacts that are irreversible and are perpetual.

50 The impact level is used to determine the requirements for peer review, notice, human involvement in decision-making and amount of training required for human operators, among other things. For example, peer review is not required for Level I decisions, peer review by at least one specified person is required for Level II and III decisions, and peer review by at least two specified persons (or publication of the specifications of the automated system in a peer-reviewed journal) is required for Level IV decisions. In addition,

73 Canada Treasury Board Directive on Automated Decision Making, Appendix B, at pp 10–11 <https://docs.google.com/document/d/1LdciG-UYeokx3U7ZzRng3u4T3IHrBXXk9JddjjueQok/edit> (accessed 17 November 2018).

Level I and II decisions may be rendered without direct human involvement, whereas Level III and IV decisions cannot be made without having specific human intervention during the decision-making process. The final decision must also be made by a human.[74]

51 Further safeguards within the Directive are that appropriate processes must be developed for all automated systems (regardless of the risk level of the decision) before an automated system goes into production to test for unintended data biases and other factors that may unfairly impact outcomes. Subsequently, routine testing must be done to ensure that the data used by a system is still relevant, accurate and up to date. The outcomes generated by all automated systems must also be monitored on an ongoing basis to safeguard against unintended outcomes and to ensure compliance with the relevant legislation and the Directive.[75]

52 Like its Canadian counterpart, the Model Artificial Intelligence Governance Framework ("Model Framework"), published by Singapore's Personal Data Protection Commission in January 2019, proposes a matrix to classify the probability and severity of harm to an individual as a result of a decision, and using that matrix to determine the level of human oversight in a decision-making process involving automated systems.[76] While developed primarily for private sector organisations, the three broad decision-making models recommended by the Model Framework are nevertheless relevant for government decisions. On one hand, a "human-out-of-the-loop" model, which involves no human oversight over the executions of decisions and no option of human override, may be used for decisions with low severity of harm. On the other hand, a "human-over-the-loop" model, which allows humans to adjust parameters during the execution of the algorithm, or a "human-in-the-loop" model, which involves the human retaining full control of a decision and the AI only providing recommendations or input, may be more suitable for decisions with higher severity of harm.

53 In the authors' view, the calibrated approaches used in Canada and suggested by the Model Framework are preferable to imposing a blanket limitation on the types of decisions that can be made by automated systems, especially machine learning systems. The risk of

74 Canada Treasury Board Directive on Automated Decision Making, Appendix C, at pp 12–13 <https://docs.google.com/document/d/1LdciG-UYeokx3U7ZzRng3u4T3IHrBXXk9JddjjueQok/edit> (accessed 17 November 2018).

75 Canada Treasury Board Directive on Automated Decision Making, Appendix C, at p 14 <https://docs.google.com/document/d/1LdciG-UYeokx3U7ZzRng3u4T3IHrBXXk9JddjjueQok/edit> (accessed 17 November 2018).

76 Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019) at paras 3.11 and 3.12.

algorithmic bias and consequent infringement of the rule on relevant considerations is minimised by testing all automated systems for bias before the production stage, and similar routine testing at the deployment stage. Although algorithmic bias cannot be completely eliminated at present,[77] human involvement in decisions with higher impact provides a second layer of safeguards against such risk. This is a balanced approach that seeks to maximise the utility of automated systems while addressing administrative law concerns.

C. *Irrationality: Specify factors considered in decision*

54 A possible solution to overcome rationality concerns is for automated systems to specify factors considered in their decisions. Take, for example, s 2(1) of the Employment of Foreign Manpower (Work Passes) Regulations 2012,[78] which enables the Controller of Work Passes to issue several different types of work passes ranging from work permits to employment passes and work holiday passes. In the hypothetical scenario where the issuance of work passes is delegated to an automated system, the automated system could specify in its decision the factors taken into consideration for each type of work pass. For example, the factors considered when issuing a work permit to a foreign employee whose occupation is "domestic worker" could be the country of origin, gender, education qualifications and general health. When issuing a work permit to a foreign employee whose occupation is "construction worker", additional factors such as experience in the construction industry and the absence of specific medical conditions may also be considered. Once such factors are specified and it is evident that they are relevant to both the statutory scheme and the decision made, the court is unlikely to consider the decision irrational even if the decision-making process or weight given to each factor may not be fully explainable. This is because these factors are strong indicators that the decision was not made arbitrarily.

---

77 Ongoing efforts are being made to develop techniques to reduce algorithmic bias. An example is a process used by a major bank to build a model that can predict whether a mortgage customer was about to refinance, with the objective of making a direct offer to the customer to retain the business. The bank started with a simple machine learning model and tested its ability to make accurate predictions, before creating more sophisticated models and testing them against the accuracy of the earlier model. By confirming that the new models were more accurate than the earlier model, the bank was able to satisfy itself that the machine learning approach was not propagating unintended biases. The bank was also able to verify that the machine learning system could achieve a balance between transparency and sophistication that was in line with the regulation framework of the financial services industry: see Chris DeBrusk, "The Risk of Machine-Learning Bias (and How to Prevent It)" *MIT Sloan Management Review* (26 March 2018).

78 S 569/2012.

55 An objection to this solution is that it is administratively inexpedient, for it requires the automated system to specify the factors considered in individual scenarios within complex administrative schemes that can generate any number of factual scenarios. On the other hand, it may be argued that specifying factors is less onerous than having to detail reasons, which administrative decision-makers are currently under no general duty to give.[79] A requirement to specify factors also avoids other difficulties engendered by a general reason-giving duty, such as the need to weigh competing interests such as national security and articulate value judgments that even human decision-makers may be unable to express.[80] In the authors' view, this solution provides a good balance between the need for rationality on the one hand, and the need for administrative efficiency on the other. This solution is not, however, a panacea to all rationality issues engendered by automated systems. It should be used in conjunction with other solutions such as requiring impact assessment, restricting automated systems from making certain categories of decisions at least until methods to ensure reliability are more developed, and instituting appropriate levels of human intervention and review.

D. Irrationality: Disclosing use of automated system and explaining algorithm used

56 Another possible solution to address rationality concerns is to disclose one or more of the following:

(a) the fact that a decision is made by an automated system ("first type of disclosure");

(b) the procedures used in an automated system, such as the types of data and AI model used ("second type of disclosure"); and

(c) the algorithm by which the decision is made ("third type of disclosure").

(1) First form of disclosure

57 The first form of disclosure does not enhance the rationality of a decision of an automated system because it reveals nothing about the decision-making process or the basis for the decision. However, it may build public confidence in the Government's use of automated systems and lead to greater acceptance of automated systems. Other related

79 *Manjit Singh s/o Kirpal Singh v Attorney-General* [2013] 2 SLR 844 at [85]; *Manjit Singh s/o Kirpal Singh v Attorney-General* [2013] 4 SLR 483 at [10].

80 *R v Higher Education Funding Council, ex parte Institute of Dental Surgery* [1994] 1 WLR 242 at 256–257.

forms of disclosure identified in the Model Framework include information on how an automated system is used in decision-making, the role and extent of an automated system in the decision-making process, and the manner in which a decision by the automated system may affect an individual.[81] The first type of disclosure is required by Art 15(1)(h) of the European Union's General Data Protection Regulation,[82] which gives a data subject the right to obtain the following information from a data controller:[83]

(a) the existence of automated systems that process the data subject's personal data to make decisions with legal or similarly significant effects on the data subject;

(b) meaningful information about the logic involved in such automated systems; and

(c) the significance and the envisaged consequences of such processing for the data subject.

58 Two related provisions are Arts 13(2)(f) and 14(2)(g), which have the same wording as Art 15(1)(h) but impose duties on data controllers to notify data subjects of the above information. Article 13(2)(f) applies where the data controller collects personal data from the data subject, whereas Art 14(2)(g) applies where the data controller obtains personal data of the data subject from another person.

(2) Second and third forms of disclosure

59 Compared to the first form of disclosure, the second and third forms of disclosure have a more direct impact on the rationality of particular decisions because they may enable the court (assisted by expert witnesses) to better understand the decision-making process and logic of a machine learning system, which may in turn support a finding that the machine learning system had not made a decision irrationally. However, there are a number of challenges to and limitations of such disclosure:

(a) First, it may be difficult to have auditing mechanisms to keep track of what goes into an automated system that is trained

81 Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019) at paras 3.28 and 3.29.

82 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (hereinafter "General Data Protection Regulation").

83 General Data Protection Regulation Art 15(1)(h) read with Arts 22(1) and 22(4).

using multiple public and private datasets.[84] One method to overcome this challenge is to put in place good data accountability practices. These include understanding the lineage of data, namely, where the data originated from, how it was collected and modified, and how its accuracy was maintained over time. Keeping a data provenance record would also enable data to be attributed and its quality ascertained based on its origin and subsequent transformation, and potential sources of error to be traced.[85]

(b) Second, there needs to be criteria to determine when a decision-maker should only explain the procedures used in an automated system, and when the algorithms that the automated system uses to make decisions must also be explained.[86] Otherwise, it would be difficult to achieve balance between the need for rationality and competing considerations such as confidentiality.

(c) Third, different machine learning algorithms are more and less easy to explain. For example, the class of algorithms known as "decision trees" are the most explainable of all the machine learning techniques because one can follow the progression of branches to determine the exact factors used in making a final decision. On the other hand, the class of algorithms known as "neural networks" are the least explainable because neural networks have hidden layers of nodes where information is transferred based on the node's activation, and each hidden node represents a non-linear combination of all the previous nodes.[87] While an alternative is to release the algorithm's source code instead of the algorithm, doing so may make simpler algorithms vulnerable to gaming and contribute little to the transparency of more complex ones. This is because the logic of many machine learning algorithms, in particular deep learning algorithms, is mostly learned from training data and rarely reflected in the source code.[88]

84 Kathryn Hume, "When is it Important for an Algorithm to Explain Itself?" *Harvard Business Review* (6 July 2018) at p 3.

85 Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019) at para 3.16a.

86 Kathryn Hume, "When is it Important for an Algorithm to Explain Itself?" *Harvard Business Review* (6 July 2018) at p 3.

87 PricewaterhouseCoopers, "Explainable AI – Driving Business Value through Greater Understanding" (2018) at pp 24–25 <https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf> (accessed 10 February 2019).

88 Kartik Hosanagar & Vivian Jair, "We Need Transparency in Algorithms, but Too Much Can Backfire" *Harvard Business Review* (23 July 2018) at p 4.

60 The above shows that the second and third forms of disclosure may not be possible in respect of certain machine learning systems. An emerging solution may lie in the development of "explainable AI" ("XAI"), which are machine learning applications that aim to give human users a degree of qualitative and functional understanding of decisions by machine learning systems.[89] Such forms of understanding may help to address rationality concerns if they are sufficient for the courts to approximate where a decision lies on the spectrum of reasonableness. XAI works by analysing the inputs used by a decision-making algorithm, measuring the impact of each input individually and in groups, and reporting the sets of inputs that had the most impact on the decision. Where XAI is applied to an essay-grading algorithm, for example, it may analyse how changes in various inputs such as content, vocabulary and grammar affected the final grade, and provide an explanation that breaks down the percentage that each input contributed to the final grade.[90] The exact degree of explainability in every case depends on factors like the complexity of the AI model in question and the explanation technique chosen. Trade-offs between the performance of a machine learning system and the explainability of its decisions are also inevitable, which necessitates a framework to consider when and the extent to which explanations are needed. From a White Paper published by PricewaterhouseCoopers, possible considerations for inclusion in such a framework include the following:[91]

(a) the utility of understanding how a machine learning system makes decisions;

(b) the number of decisions that the machine learning system has to make;

(c) the potential impact of a decision by the machine learning system; and

(d) the level of human involvement in a decision by the machine learning system.

61 While it is too early to tell whether XAI can solve the "black box" problem, its objective of seeking to provide the optimum level of transparency while not necessarily requiring huge amounts of possibly

89 PricewaterhouseCoopers, "Explainable AI – Driving Business Value through Greater Understanding" (2018) at p 4 <https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf> (accessed 10 February 2019).

90 Kartik Hosanagar & Vivian Jair, "We Need Transparency in Algorithms, but Too Much Can Backfire" *Harvard Business Review* (23 July 2018) at p 5.

91 PricewaterhouseCoopers, "Explainable AI – Driving Business Value through Greater Understanding" (2018) at pp 6–7 <https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf> (accessed 10 February 2019).

unexplainable data to be disclosed and explained[92] is worth pursuing. Recent studies have also shown that XAI is within the capabilities of today's machine learning and statistical methods.[93]

E. Irrationality: Repeatability as interim alternative to explainability

62 Where it is not possible to achieve explainability with the current state of technology, the Model Framework suggests that "repeatability" could be an interim, albeit not equivalent, alternative. The Model Framework defines "repeatability" as "the ability to consistently perform an action or make a decision, given the same scenario".[94] Practices that promote repeatability include conducting repeatability assessments in live environments, performing counterfactual fairness testing, and assessing how exceptions can be identified and handled when repeatability cannot be achieved.[95]

63 It is suggested that repeatability of itself is unlikely to satisfy the requirements of rationality because, like the first form of disclosure, it reveals nothing about the decision-making process or basis for a decision of an automated system. Further, it is possible for an automated system to make irrational decisions in a consistent manner. Nevertheless, repeatability may be used to show that decisions of an automated system are not arbitrarily made, and that the same criteria and reasoning (even if not explainable) are consistently applied to persons with the same relevant attributes. In so doing, repeatability provides a degree of transparency and may build public confidence in the use of automated systems.

F. Improper delegation or devolution: Enable human decision-maker to retain control or provide express legislative authorisation

64 At the outset, a distinction must be made between a decision that is completely automated and a decision that contains input generated by an automated system. Where only part of the decision-making process is automated and the decision-maker retains control

92 Kartik Hosanagar & Vivian Jair, "We Need Transparency in Algorithms, but Too Much Can Backfire" *Harvard Business Review* (23 July 2018) at p 4.
93 Kartik Hosanagar & Vivian Jair, "We Need Transparency in Algorithms, but Too Much Can Backfire" *Harvard Business Review* (23 July 2018) at pp 4–5.
94 Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019) at para 3.22.
95 Personal Data Protection Commission, *A Proposed Model Artificial Intelligence Governance Framework* (January 2019) at para 3.22.

over the outcome, there is arguably no delegation, and objections on grounds that a power was not properly delegated to the automated system do not apply. Section 29(1) of the Interpretation Act, which provides that a power conferred also includes incidental powers that are "reasonably necessary to enable the person to do or enforce the doing of the act or thing", is arguably wide enough to enable a decision-maker to enlist the help of an automated system without express statutory authorisation. This is in theory no different from a decision-maker enlisting the help of existing technological tools or subordinate officers to assist in parts of the decision-making process.[96]

65 On the other hand, express legislative authorisation is arguably required to enable the relevant power to be delegated. This will require either amendments to s 36(1) of the Interpretation Act to enable power to be delegated to such systems, or amendments to individual legislation to enable delegation to automated systems under specific regulatory regimes. An example of such express legislative authorisation is s 2(1) of the UK's Social Security Act 1998,[97] which enables certain powers of the Secretary of State to be made by "by an officer of his acting under his authority" or "by a computer for whose operation such an officer is responsible".

66 As for devolution using the *Carltona* principle, a distinction must again be made between a decision that is completely automated and a decision that contains input generated by an automated system. Where a devolvee retains control over the final decision, the analysis above[98] applies and the issue of whether a power has been devolved to the automated system does not arise. Where a decision is completely automated, devolution of the relevant power to an automated system is arguably permissible as long as the two grounds of administrative necessity and ministerial responsibility are met. This is because the issue of who is a "responsible official" to whom ministerial powers may be delegated has always been an issue for the Minister to decide, and it is the Minister who must account to Parliament for his choice.[99]

67 In both cases, there arguably remains a more fundamental issue of whether the concepts of delegation and devolution can be

96 See also Cary Coglianese & David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era" (2017) 105 Geo LJ 1147 at 1183.

97 c 14.

98 See para 64 above.

99 See *Carltona Ltd v Commissioners of Works* [1943] 2 All ER 560 at 563, where the English Court of Appeal held that:

> … if for an important matter [the Minister] selected an official of such junior standing that he could not be expected competently to perform the work, the Minister would have to answer for that in Parliament.

appropriately used in the context of automated systems. This is because it is arguable that unlike human delegates and devolvees, an automated system (being a computer program) can never truly be said to act independently of its designer or the relevant government agency.[100] While legislative techniques have been developed to deem a decision made by an automated system to be a decision of the human decision-maker, such deeming provisions have been criticised as being "highly artificial constructs of decision-making processes".[101] It may therefore be necessary to develop more sophisticated techniques to resolve this issue satisfactorily.

V. Conclusion

68 The "Digital Government Blueprint", which is a statement of the Government's ambition to leverage data and harness new technologies to deliver better public services, was launched in June 2018. The overarching objective is to create: "A Government that is Digital to the Core, and Serves with Heart". Being "digital" means to use data, connectivity and computing decisively to transform the way the Government serves citizens and businesses, and the way public officers are enabled to contribute fully to their work. "Serving with Heart" means automating work wherever possible in order to provide a personal touch in a way that enriches the experience of the engagement between the Government and citizens.[102] Among other things, the Government has announced that it will identify high-impact areas for the deployment of AI in government, including automating rule-based tasks, providing personalised and anticipatory services, and anticipating situations such as traffic or security incidents.[103] Public officers will also be supported by data and automation to execute high quality decisions and processes in a timely manner.[104]

100 Melissa Perry, "iDecide: Administrative Decision-Making in the Digital World" (2017) 91 ALJ 29 at 31.
101 Melissa Perry, "iDecide: Administrative Decision-Making in the Digital World" (2017) 91 ALJ 29 at 31.
102 Smart Nation and Digital Government Office, "Digital Government Blueprint" (June 2018) at p 5 <https://www-smartnation-sg-admin.cwp.sg/docs/default-source/default-document-library/dgb_booklet_june2018.pdf> (accessed 21 November 2018).
103 Smart Nation and Digital Government Office, "Digital Government Blueprint" (June 2018) at p 16 <https://www-smartnation-sg-admin.cwp.sg/docs/default-source/default-document-library/dgb_booklet_june2018.pdf> (accessed 21 November 2018).
104 Smart Nation and Digital Government Office, "Digital Government Blueprint" (June 2018) at p 9 <https://www-smartnation-sg-admin.cwp.sg/docs/default-source/default-document-library/dgb_booklet_june2018.pdf> (accessed 21 November 2018).

69 It is clear from the above that automated decision-making is on its way and will be an essential part of the "Digital Government" of the future. Granted, the administrative law issues engendered by automated systems necessitate careful consideration and appropriate safeguards when agencies adopt such systems. Case-by-case assessments are likely to be necessary, and the result could well be that automated systems prove to be promising in some instances and less advantageous in others. However, this does not mean that society should eschew the use of automated systems, which are already widely used in the private sector,[105] in government agencies. Instead, a concerted effort needs to be made on identifying the types of administrative decisions to be automated, and designing automated systems with the appropriate safeguards in place. By suggesting possible solutions, this article has sought to show that the administrative law issues surrounding automated decision-making in government agencies are not insurmountable.

70 The legal profession and public law scholars have a valuable role to play in making the vision of a "Digital Government" a reality. A recent initiative is the launch of the Centre for AI and Data Governance at the Singapore Management University, which seeks to "promote cutting-edge thinking and practices in AI, and data policies and regulations".[106] Its current research projects include studying the legal, ethical and social dimensions of using big data, examining the opportunities and challenges AI and big data present to the financial system, and examining how existing intellectual property norms will be challenged by developments in AI. In the administrative law space, much work is needed to identify issues specific to different types of automated systems that are developed in the future, and offering analysis and solutions to government agencies that seek to implement such systems. The case studies in this article also highlight the importance of involving lawyers in the design and review of automated systems so that they incorporate principles of administrative law by design, and continue to do so as they evolve. With the active involvement and support of the legal community, there is much room for optimism that automated systems can remain consistent with the principles of administrative justice while enhancing public service delivery.

---

105 Cary Coglianese & David Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-Learning Era" (2017) 105 Geo LJ 1147 at 1160.

106 Singapore Management University, "SMU Leverages Multi-Disciplinary Expertise, Launches Centre for AI and Data Governance" (24 September 2018) <https://www.smu.edu.sg/news/2018/09/24/smu-leverages-multi-disciplinary-expertise-launches-centre-ai-and-data-governance> (accessed 25 April 2019).