Reason-Giving in the Age of Algorithms

JESSICA PALAIRET*

Public sector agencies are increasingly using sophisticated machine learning algorithms to assist and make decisions previously made solely by humans. This use of advanced algorithms presents opportunities, but also great risks to administrative law. Many complex algorithms are "black boxes", meaning no person can explain how they work. Further, "dirty data" and machine code can produce discriminatory or biased decisions that are difficult to identify and regulate. The idea that keeping a "human in the loop" will address these problems is unrealistic and short-sighted. Therefore, this article argues that the use of AI in administrative decision-making necessitates the development of a general duty on decision-makers to provide reasons for their decisions. Administrative law cannot stand still amidst the rise of artificial intelligence. A duty to give reasons is not a silver bullet solution, but it is an essential response to decision-making in the age of algorithms.

I INTRODUCTION

Over the past 10 years, the use and sophistication of machine learning algorithms have increased at a remarkable rate. We are in the midst of a technological revolution that could either be society's greatest opportunity or its most pressing threat.[1] New Zealand's public sector is not immune to this revolution. Public sector agencies are increasingly using artificial intelligence (AI) to make decisions previously made by humans. These decisions range from the automatic approval of Accident Compensation Corporation (ACC) claims to decisions on welfare grants, school funding and immigration.[2]

These applications of AI are just the beginning. This article focuses on machine learning algorithms that derive rules and predictions from large swathes of data. These algorithms can be used to support human decision-

* BA/LLB(Hons), University of Auckland. The author would like to thank Professor Janet McLean QC for her supervision and support, and Matt Bartlett for his help and inspiration.
This article was awarded the MinterEllisonRuddWatts Writing Prize for 2020.

1 Yuval Noah Harari "Why Technology Favors Tyranny" *The Atlantic* (online ed, New York, October 2018).

2 Stats NZ *Algorithm Assessment Report* (October 2018) at 36–38.

makers, recommend decisions and outcomes for humans to approve, or make decisions on their own, eliminating humans from the decision-making process altogether. Algorithmic decision-making presents an important tension. On the one hand, these technologies might be transformative, increasing the efficiency, consistency and cost-effectiveness of decision-making. On the other, the opacity of AI challenges administrative law's ability to hold decision-makers accountable.

There is currently no general duty on decision-makers to give reasons for their decisions. This article argues that the use of AI in public sector decision-making necessitates the development of such a duty. Without any requirement for reasons, the ability of individuals and courts to understand how an AI makes decisions, and assess whether or not those decisions are fair, is compromised.

Parts II and III of this article are contextual. Part II provides necessary background on the technology and its application in New Zealand public sector agencies. Part III outlines the state of the law on the duty to give reasons. It argues that under the existing legal framework, there are no adequate ways of ensuring the transparency of AI tools in the public sector. In the absence of this duty, Part IV presents three key challenges algorithmic decision-making poses to administrative law. First, the increasing complexity of algorithms turns them into "black boxes", meaning no human can understand how they produce their decisions. Secondly, algorithmic tools heighten the risk of discrimination and bias in ways that are incredibly difficult to identify. Finally, the idea of "keeping a human in the loop", as currently suggested by the government, risks unduly fettering decision-makers' discretion and provides insufficient protection against potential mistakes.

Part V focuses on solutions. It outlines what this general duty might look like at a high level. Individuals should have the right to be provided with reasons in cases where algorithms have assisted or made decisions that affect them. This article recommends imposing a corresponding obligation on the developers of algorithmic tools used in the public sector to create "explainable" AI systems. At a macro level, it also proposes the development of a watchdog agency to oversee the deployment of AI in the public sector. Finally, this article details why the development of a duty is necessary, with reference to the importance of incentives and reason-giving to the democratic legitimacy of administrative law.

II ALGORITHMS IN THE PUBLIC SECTOR

At the most basic level, algorithms are procedures or formulae used to solve a problem or carry out a task.[3] However, with the rise of big data and more advanced technologies, the complexity and capabilities of algorithms have increased.[4] Although there is a range of definitions and types of algorithms, this article focuses on a subset of advanced machine learning algorithms. Specifically, it focuses on the subset that uses analytical processes to interpret information resulting in, or materially informing, decisions that impact individuals or groups.[5] Machine learning gives algorithms the ability to predict likely outcomes based on historical data sets.[6] These data sets are becoming increasingly vast.[7] The idea is that the more data available to AI, the more accurate its results will be.[8] This accuracy is possible because machine learning algorithms learn directly from data, identifying patterns that produce meaningful information about inputs. Put another way, algorithms are no longer just executing pre-written instructions. They are arriving at solutions to problems based on patterns in data that humans may not be able to even perceive.

While there are several different approaches to machine learning, one of the most common involves the use of artificial neural networks.[9] Artificial neural networks were first conceived and adopted in the late 1940s, but the last decade has seen a number of significant breakthroughs. These breakthroughs have produced improvements in performance, especially where multiple layers of neural networks are involved.[10] Neural networks are based on the fundamental operating principles of the human brain. The human brain contains as many as 100 billion neuron cells that process information.[11] Neural networks simulate connections between neurons in the human brain to learn and adapt from large amounts of data.[12] Several layers of mathematically simulated neurons work together to find patterns and make connections between data points. As the network processes data, its accuracy improves. The network is constantly self-learning and self-optimising as it works through data, without the need for human control.

---

3 At 5.

4 Max Tegmark *Life 3.0: Being Human in the Age of Artificial Intelligence* (Allen Lane, London, 2017) at 103.

5 This is in line with the definition used in Stats NZ, above n 2, at 5.

6 At 8.

7 New Zealand Data Futures Forum *New Zealand's Data Future* (2014) at 15.

8 Tegmark, above n 4, at 128.

9 Martin Ford *The Rise of the Robots: Technology and the Threat of Mass Unemployment* (Oneworld Publications, London, 2016) at 92–93; and Yavar Bathaee "The Artificial Intelligence Black Box and the Failure of Intent and Causation" (2018) 31 Harv JL & Tech 889 at 901–903.

10 Ford, above n 9, at 94.

11 At 93.

12 Tegmark, above n 4, at 148.

In October 2018, Internal Affairs and Statistics New Zealand published the *Algorithm Assessment Report*, the first stocktake on the use of algorithms in the New Zealand public sector.[13] It identified 33 uses of algorithms across 14 public sector agencies, many of which may be subject to judicial review.[14] The Ministry of Education, for example, uses algorithms to allocate resources and make other operational decisions that impact children.[15] The data collected includes information about the age, gender and ethnicity of students and details about their attendance, performance, discipline and engagement.[16] Oranga Tamariki also collects and uses a wide range of information about vulnerable children to inform decisions on matching children with caregivers, and manage the caseload of frontline staff.[17] In 2017, the agency received 158,900 referrals, from which it was determined 33,000 children and young people required further assistance.[18] These examples highlight the impact and reach of algorithmically assisted decisions.

One of the most striking uses of AI in the public sector is in the Young People Not in Employment, Education or Training (NEET) programme for school leavers. The AI evaluates demographic information and factors concerning the young person. These factors include schooling history, whether a young person's parents have been on a benefit, and whether the young person has been the subject of a notification to Oranga Tamariki.[19] An algorithm then identifies which school leavers are at a high, medium or low risk of long-term unemployment.[20] This risk indicator rating is given to NEET providers who make contact and offer assistance. Over 60,000 young people have accepted this assistance since 2012.[21] Algorithms are also at the heart of the government's "wellbeing" approach.[22] In essence, the government uses algorithms, such as the Treasury's CBAx tool, to make high-level budget decisions based on predictions about individuals and groups in society.[23] This approach places predictive algorithmic analysis at the centre of government decision-making and provides important context for the use of algorithmic tools in the public sector.

Finally, it is important to consider the future of the use of algorithms in public sector decision-making. It is likely that the use of algorithms will

13 Stats NZ, above n 2.
14 At 9 and 36–38.
15 At 13.
16 At 13.
17 At 13.
18 At 13.
19 At 14.
20 At 14.
21 At 14.
22 For a description of the wellbeing framework, see New Zealand Treasury "Our living standards framework" (12 December 2019) <www.treasury.govt.nz>.
23 New Zealand Treasury "The Treasury's CBAx Tool" (30 September 2019) <www.treasury.govt.nz>.

only increase. The economic advantages of using algorithms are compelling.[24] Algorithmic decision-making can be substantially faster and less expensive than human labour, and its quality improves over time.[25] Unlike humans, AI does not demand a wage, does not tire and is not limited to the hours of the working day. Some also claim AI will increase the impartiality of decision-making, as it is objective in ways humans are not.[26] Importantly, AI might also become better at making some decisions than humans.[27] According to the median estimate of experts, there is a 50 per cent chance that AI will reach a higher level of general intelligence than humans in the next 30 years.[28] This chance rises to 90 per cent within 60 years.[29] This article is not focused on the applications of future super-intelligent AI. However, the importance of developing explainable AI is heightened when considering its future impact. While ensuring transparency in decision-making tools is important now, it will be imperative moving forward.

III THE STATE OF THE LAW ON GIVING REASONS

Administrative law is fundamentally concerned with the court's constitutional responsibility to uphold the rule of law.[30] Judicial review helps ensure public officials act within the law and are held accountable if they do not.[31] An important aspect of the rule of law is transparency.[32] At a high level, transparency relates to the need for the government to be open and accountable in respect of its rules and decisions. In the context of administrative law, a central tenet of transparency is that decisions are not made in "smoke-filled back rooms".[33] People should be able to understand the reasons for decisions that affect them. Transparency in this sense ensures the accountability of public officials who exercise discretionary powers, while safeguarding against the abuse of these powers.

---

24 Lee Rainie and Janna Anderson "Theme 2: Good things lie ahead" (8 February 2017) Pew Research Center <www.pewresearch.org>.

25 Tegmark, above n 4, at 13–39.

26 Stats NZ, above n 2, at 26–27.

27 See generally Nick Bostrom *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press, Oxford, 2014).

28 Vincent C Müller and Nick Bostrom "Future Progress in Artificial Intelligence: A Survey of Expert Opinion" in Vincent C Müller (ed) *Fundamental Issues of Artificial Intelligence* (Springer Nature, Switzerland, 2016) 555 at 563–564.

29 At 563–565.

30 *Tannadyce Investments Ltd v Commissioner of Inland Revenue* [2011] NZSC 158, [2012] 2 NZLR 153 at [3].

31 Matthew Smith *The New Zealand Judicial Review Handbook* (2nd ed, Thomson Reuters, Wellington, 2016) at 7.

32 Monika Zalnieriute, Lyria Bennett Moses and George Williams "The Rule of Law and Automation of Government Decision-Making" (2019) 82 MLR 425 at 429–430.

33 *Hamilton City Council v Waikato Electricity Authority* [1994] 1 NZLR 741 (HC) at 746.

Transparency takes many forms, but this article focuses on giving reasons for decisions.[34] Giving reasons serves a range of instrumental and non-instrumental benefits.[35] In addition to its role in upholding the rule of law, openness is key to maintaining public confidence in administrative decision-making.[36] People ought to be able to understand why decisions about them are made. Reasons ensure justice is done, and is seen to be done. Giving reasons has also been explained as an important ingredient in the overall fairness of decision-making. It serves a "dignitarian" function:[37] as Jerry Mashaw contends, "authority without reason is literally dehumanising".[38] Placing human dignity and freedom at the centre of the rationale for a duty to give reasons is consistent with New Zealand's international human rights obligations and the principles underpinning them.[39] From a practical standpoint, reason-giving incentivises good decision-making. It disciplines decision-makers, helping direct their minds towards the right questions and encourages a thorough approach to each decision they make.[40] Reasons also help both the person about whom a decision is made and the court to assess the lawfulness of a decision. Finally, reasons may result in more efficient use of court resources, as more meritorious claims are brought forward and less meritorious claims are not.

However, there is no general common law duty to provide reasons for administrative decisions. This absence has been justified on several grounds. Reason-giving might not always be necessary or desirable, particularly where giving reasons will increase the administrative burdens on decision-makers.[41] There are some instances where giving reasons would threaten important interests, such as national security. Further, there is a risk that some decision-makers will make worse decisions in the public gaze.[42] There might be pressure to make "easy" decisions, but not necessarily the right ones. A requirement to give reasons might also lead to generalised decisions less focussed on the individual circumstances of each case, resulting in box-checking exercises. Relatedly, it might demand the

34 Jens Forssbæck and Lars Oxelheim "The Multifaceted Concept of Transparency" in Jens Forssbæck and Lars Oxelheim (eds) *The Oxford Handbook of Economic and Institutional Transparency* (Oxford University Press, New York, 2014) 3 at 4.

35 PP Craig "The Common Law, Reasons and Administrative Justice" (1994) 53 CLJ 282 at 283.

36 At 283.

37 TRS Allan "Procedural Fairness and the Duty of Respect" (1998) 18 OJLS 497 at 499.

38 Jerry L Mashaw "Public Reason and Administrative Legitimacy" in John Bell and others (eds) *Public Law Adjudication in Common Law Systems: Process and Substance* (Hart Publishing, Oxford, 2016) 11 at 17.

39 See, for example, *Universal Declaration of Human Rights* GA Res 217A (1948), art 1; Tim Cochrane "A General Public Law Duty to Provide Reasons: Why New Zealand Should Follow the Irish Supreme Court" (2013) 11 NZJPIL 517 at 533–534; and Sian Elias "Administrative Law for 'Living People'" (2009) 68 CLJ 47 at 65.

40 *Lewis v Wilson & Horton Ltd* [2000] 3 NZLR 546 (CA) at [82].

41 Mark Elliott "Has the Common Law Duty to Give Reasons Come of Age Yet?" [2011] PL 56 at 64.

42 Frederick Schauer "Transparency in Three Dimensions" [2011] U Ill L Rev 1339 at 1349.

appearance of unanimity when there is diversity.[43] Finally, a general duty might create problems in cases where officials had relied upon value judgments not capable of easy expression.[44] These reasons may fail to show why a decision was reached, and so may not be effective in any event.

Many of these arguments fall away when considered in the context of algorithmic decision-making. Algorithms are immune to political and popular pressure. The incentives that apply to people in decision-making contexts simply do not apply to algorithmic code. Further, one of the main benefits of AI is speed. Where bureaucrats may take a long time to make decisions, AI can make decisions significantly faster, which is advantageous to the general public. It may also be possible to code in a requirement for reasons for a decision to be provided as an output.[45] This could simplify the provision of reasons, again not requiring additional time.

While there is no general common law duty to give reasons, courts have developed limited instances where reasons are required. These include where the right or interest at stake is fundamental; an interest "so highly regarded by the law (for example, personal liberty), that fairness requires that reasons, at least for particular decisions, be given as of right".[46] A further instance is "where there is some 'trigger factor' specific to the case in question".[47] An example of this might be where a decision appears inexplicable, as was seen in *R v Civil Service Appeal Board, ex parte Cunningham*.[48] The Board's decision that the claimant, a prison officer, had been unfairly dismissed and should receive compensation, was considered "aberrant".[49] The meagre award of compensation, Leggatt LJ said, "looks as though [it] is less than it should be, and yet he has not been told the basis of assessment".[50]

Another category is where decisions depart from existing policy, as claimants have a legitimate expectation that decision-makers will follow policy. A version of this exception was applied in *Regina v Secretary of State for Trade and Industry, ex parte Lonrho plc*.[51] There, the Court held a lack of reasons may justify an inference that the decision was incorrect where "all other known facts and circumstances appear to point overwhelmingly in favour of a different decision".[52]

---

43 *Regina v Higher Education Funding Council, ex parte Institute of Dental Surgery* [1994] 1 WLR 242 (QB) at 256–257.

44 Matthew Groves "Before the High Court: Reviewing Reasons for Administrative Decisions: *Wingfoot Australia Partners Pty Ltd v Kocak*" (2013) 35 Syd LR 627 at 633.

45 Colin Gavaghan and others *Government Use of Artificial Intelligence in New Zealand* (New Zealand Law Foundation, 2019) at 43.

46 *Dental Surgery*, above n 43, at 263.

47 Elliott, above n 41, at 58.

48 *R v Civil Service Appeal Board, ex parte Cunningham* [1991] 4 All ER 310 (CA).

49 At 325.

50 At 325.

51 *Regina v Secretary of State for Trade and Industry, ex parte Lonrho plc* [1989] 1 WLR 525 (HL).

52 At 539–540.

These exceptions cover wide ground. Elias LJ has noted:[53]

> … it may be more accurate to say that the common law is moving to the position whilst there is no universal obligation to give reasons in all circumstances, in general they should be given unless there is a proper justification for not doing so.

In *Lewis v Wilson & Horton*, the Court of Appeal indicated it would like to re-consider whether there should be a general rule requiring judges to give reasons at "an early opportunity".[54] However, almost 20 years on, that opportunity has not arisen. New Zealand law has only developed in some discrete areas since *Lewis*. Much of New Zealand case law on the duty to give reasons concerns the courts' duty to give reasons.[55] For example, in *MA v Legal Services Agency*, a duty was rejected in the context of legal aid applications. The Court concluded that the Legal Aid Agency's obligation was "oblique" and only imposed under s 23 of the Official Information Act 1982 (OIA).[56] Despite the Court's statement in *Lewis*, there have been no advances towards a general common law duty of any sort.

Based on common law, many algorithmic decisions will not need to be accompanied by reasons. Some algorithmic decisions will likely fall under the exceptions to the common law rule, such as where a decision impacts a fundamental right or is clearly "aberrant". However, Sedley J's reasoning in *R v Higher Education Funding Council, ex parte Institute of Dental Surgery* indicates how high the bar is for reading in a common law duty on that basis.[57] In that case, the Institute of Dental Surgery sought to review a decision of the Universities Funding Council to downgrade its research grading, substantially reducing the research funds available. No reasons were provided for that decision. Sedley J held that although the lack of reasons frustrated the Institute's rights to review, fairness alone cannot always require reasons to be given; otherwise, the duty would apply universally.[58] Whether reasons should have been provided required balancing a number of factors, including the openness of the procedure, the rights at issue, and the fact the decision was based on academic judgment.[59] This case hints at the reluctance of the courts to require reasons, even if fairness is compromised in some way.

Further, the circumstantial way by which courts determine whether reasons ought to have been provided is problematic. AI developers need to

---

53 *Regina (Oakley) v South Cambridgeshire District Council* [2017] EWCA Civ 71, [2017] 1 WLR 3765 at [30].
54 *Lewis*, above n 40, at [85].
55 See *R v Awatere* [1982] 1 NZLR 644 (CA); *Lewis*, above n 40; and *R v Taito* [2003] UKPC 15, [2003] 3 NZLR 577.
56 *MA v Legal Services Agency* HC Auckland CIV-2008-404-6803, 11 December 2009 at [39].
57 *Dental Surgery*, above n 43, at 257.
58 At 256–257 and 261.
59 At 260–261.

be able to know what rules to apply. Individuals also need to know how to apply these rules to decide whether to invest the large cost associated with challenging a decision where no reasons were provided. The lack of clarity in the common law creates obstacles in the context of AI decision-making.

The New Zealand Law Foundation (NZLF) has suggested in a recent report that a potential solution in the context of algorithmic decision-making exists in s 23 of the OIA. This section provides individuals with a right to a statement of reasons in all decisions made by an authority covered by the OIA, if that decision affects the individual in their personal capacity.[60] The right to reasons under the OIA, however, is different from a right to have reasons stated in a decision itself. Section 23 only applies when a request is made. A more comprehensive common law duty would require reasons to be provided when the decision is made, regardless of any request. This difference is essential, as it takes a determined plaintiff to request reasons. It may not be possible to identify mistakes until reasons are provided. AI developed by private companies might also hide behind exceptions to the OIA, which prevent the code from being inspected.[61] For decisions where the decision-making body is subject to neither the OIA nor a separate statutory duty, the common law provides the only obligation to provide a reasoned decision. The OIA cannot be seen as a substitute for a general common law duty.

In response to the use of algorithms in the public sector, the government has signalled its commitment to developing transparent and accountable algorithms. Statistics New Zealand has created a draft Algorithm Charter, which commits government agencies to use algorithms in a fair, ethical and transparent way.[62] It draws on principles developed by the Privacy Commissioner and the government's Chief Data Steward on the safe and effective use of data and analytics by government agencies.[63] These principles include a requirement to maintain transparency over algorithmic activities, understand the limitations of data analytics and focus on the "people behind the data, and how to protect them against misuse of information".[64]

Agencies are also independently developing frameworks around their predictive modelling algorithms. The Ministry of Social Development (MSD) has created a Privacy, Human Rights and Ethics Framework. Teams within MSD use this framework to ensure human rights and ethics are adequately considered when using predictive algorithms in projects.[65]

60 Gavaghan and others, above n 45, at 57 and 75.
61 See Official Information Act 1982, s 9.
62 Stats NZ *Algorithm Charter* (October 2019).
63 Privacy Commissioner and Stats NZ *Principles for the safe and effective use of data and analytics* (May 2018).
64 At 1.
65 Ministry of Social Development *The Privacy, Human Rights and Ethics (PHRaE) Framework* (2018).

Importantly, in May 2019, the New Zealand government signed on to the OECD's intergovernmental principles on AI.[66] One of these principles is for the government to commit to transparent and responsible disclosure regarding AI systems.[67] This principle seeks to enable those affected by AI to understand and challenge outcomes, based on plain and easily understandable information on the factors and logic that served as the basis for the prediction, recommendation or decision. Although the OECD principles are soft international law and thus non-binding,[68] they are nonetheless significant.

These developments signify the government's commitment to ensure explainable and transparent AI is used in decision-making. What the developments do not provide is a way of implementing this across government. The lack of any enforceable general duty to provide reasons means New Zealand is falling short of these objectives. While the OIA provides a right to reasons after a decision has been made, this alone is insufficient. The next Part considers three key problems that arise in administrative law when AI is used without any duty to give reasons. Together, these problems form the backbone of the case for the development of such a duty.

IV CHALLENGES ALGORITHMIC DECISION-MAKING POSES TO ADMINISTRATIVE LAW

Black Box Algorithms

A contemporary machine learning algorithm is a "black box".[69] The complexity of AI systems means it is not always possible to understand how algorithms arrive at decisions. Machine learning uses complex neural networks that mimic the functioning of the human brain. AI, therefore, derives results independent of human control. We may know an algorithm's inputs and outputs, but exactly how the AI computes its results is unclear. Consider a recent incident at Mount Sinai Hospital in New York.[70] The hospital applied a deep learning algorithm, "Deep Patient", to the hospital's database of 700,000 patient records. Deep Patient was able to discover hidden patterns in hospital data that anticipated the onset of psychiatric disorders, such as schizophrenia, surprisingly well. These disorders are

66 OECD *Recommendation of the Council on Artificial Intelligence* (OECD/Legal/0449, May 2019).
67 At [1.3].
68 OECD "Soft Law" (2019) <www.oecd.org>.
69 Frank Pasquale *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, Cambridge (Mass), 2015).
70 Ariel Bleicher "Demystifying the Black Box That Is AI" *Scientific American* (online ed, New York, 9 August 2017).

notoriously difficult for physicians to predict, and the hospital and AI programmers did not know how the AI had identified them. This situation raises some problems: can a doctor tell a patient they are likely to develop schizophrenia without having any idea why? Further, should we trust AI when we cannot understand it?

Machine learning algorithms can also produce solutions humans cannot understand. AI is often programmed to "think" differently to humans, producing new and innovative solutions to problems. A well-known example is in the game of chess. Google's DeepMind AI learned how to play chess, defeating the best chess players and existing computer programs in a matter of hours.[71] Its machine learning algorithm was able to develop a set of playing strategies that humans had never considered before. This example shows how AI can "think" differently from humans and produce alternative answers. However, the issue becomes more pronounced where the applications of AI are more complex (and important) than chess. An example is the problem of dimensionality. Humans cannot visualise high-dimensional patterns and shapes, but this may be one way machine learning algorithms break down information and reach a result.[72] Some AI models, for example, contain nearly one trillion parameters.[73] AI may, therefore, produce solutions beyond what the brain can understand.

These issues pose a significant challenge for administrative law. A key element of natural justice is the ability to understand why decisions have been made, and to have the opportunity to respond. This is particularly important in the context of administrative law, which often concerns decisions which are important to people's lives — much more so than the ability of an AI to beat a human at chess. Paul Craig notes that failing to provide reasons risks creating a "Kafkaesque world" in which decisions are made without people having the ability to understand the reasons for those decisions.[74] The immediate issue opaque algorithmic decision-making presents is that no person, including the developer of the AI, can explain how or why a decision was reached. This will frustrate an individual's right to review. If people do not know why a decision was made, they will be unable point to a particular error and courts will be unable to effectively review the algorithmic decision.

However, concerns around the black box nature of AI are not universal. John Zerilli and others argue that concerns around the inability to understand AI decision-making are overstated.[75] The authors argue that the focus on the transparency of AI decision-making systems holds AI to a

71 James Somers "How the Artificial-Intelligence Program AlphaZero Mastered Its Games" *The New Yorker* (online ed, New York, 28 December 2018).
72 Bathaee, above n 9, at 903.
73 Bleicher, above n 70.
74 Craig, above n 35, at 284.
75 John Zerilli and others "Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?" (2019) 32 Philosophy & Technology 661.

higher standard than human decision-making. Just as we may not understand how AI makes its decisions, we "don't know very much about how the brain works" either.[76] Human decision-makers are never expected to explain the cognitive processes that lead to their conclusions. As a result, Zerilli is concerned that black box AI perpetuates a "double standard" in which "machine tools must be transparent to a degree that is in some cases unattainable … while human decision-making can get by" with a lower standard of transparency.[77] This double standard might lead to a chilling effect on AI development.

The authors' criticism ignores the fundamental differences between the human brain and AI. If a human decision-maker fails to provide sufficient contemporaneous reasons for decisions, he or she will generally be able to provide ex post reasons. Scientists may not understand exactly how the neurons in the decision-maker's brain fired to reach a decision, but human decision-makers are nonetheless able to communicate intelligible reasons on request. The issue with AI is that it might never be able to communicate the reasons for its decisions in an understandable way. Further, understanding the inner workings of the human brain is not essential for showing decisions were reached fairly, but understanding the inner workings of AI sometimes will be. There might, for example, be fundamental mistakes in algorithmic code that produce aberrant results. Finally, the level and quality of reason-giving required already depend on the type of decision and adjudicator. This is why judges are subject to much more rigorous standards of reason-giving than, for example, school boards.[78] AI decision-making may fall on the more serious end of that spectrum. Requiring fairly rigorous standards of reason-giving for AI reflects the importance of the decisions it makes and the question marks over the technology. As a result, there is no issue with AI being held to a higher standard of transparency than human decision-makers. Rather, it is necessary to safeguard people's ability to understand and challenge decisions that affect them. This is fundamental.

This Part has argued that the black box nature of AI decision-making demands a reconception of the duty to provide reasons as it applies to AI. Humans may not be able to understand how or why AI reaches its decisions, regardless of whether a human is "kept in the loop". The incomprehensibility of AI outputs takes Craig's concern that a failure to furnish reasons creates a "Kafkaesque world" to another level. The next Part pays specific attention to the problem of AI as a biased decision-maker.

76 At 666.
77 At 668.
78 Smith, above n 31, at 858.

Algorithmic Biases

1 The Risk of Discriminatory Algorithms

Algorithms are products of their data.[79] Issues, therefore, arise when the data algorithms learn from is flawed — the "dirty data problem".[80] Data can be "dirty" if it is unrepresentative, contains any latent errors or reflects historical structures and assumptions containing biases that society does not wish to re-entrench.[81] Any biases or errors in underlying data fed to AI will be continually reproduced and deeply embedded within AI systems through the machine learning process. There are several relevant examples:

- Google's facial recognition app classed some black people as gorillas.[82]
- Amazon's now abandoned recruitment tool was biased against hiring women as the underlying data was based on the resumes of employees that had been previously selected for these positions — most of whom were white males.[83]
- COMPAS, a predictive risk assessment tool used widely in the United States for sentencing, flagged black individuals as a high risk to society almost twice as often as white individuals.[84]

These examples show that overrepresentation of a minority in a dataset might distort predictive risk assessments. Algorithmic decisions will always reflect the quality and inclusiveness of the data inputs provided.

This risk exists even where agencies endeavour to avoid producing discriminatory results. The Department of Corrections uses an algorithm for calculating an offender's risk of reconviction that excludes ethnicity as a variable.[85] Corrections removed ethnicity from the algorithm even after it found that doing so decreased the overall accuracy of the tool by two per cent. The removal was presumably to ensure the algorithmic tool would not produce racist outcomes.[86] However, taking the ethnicity variable out of the

79 Ford, above n 9, at 88–91.
80 Zerilli and others, above n 75, at 672.
81 Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16 Duke Law & Technology Review 18 at 28.
82 Mauro Comi "Is Artificial Intelligence Racist? (And Other Concerns)" (12 November 2018) Towards Data Science <www.towardsdatascience.com>.
83 Jeffrey Dastin "Amazon scraps secret AI recruiting tool that showed bias against women" (10 October 2018) Reuters <www.reuters.com>.
84 Julia Angwin and others "Machine Bias" (23 May 2016) ProPublica <www.propublica.org>.
85 Stats NZ, above n 2, at 21.
86 Department of Corrections *Over-representation of Māori in the criminal justice system: An exploratory report* (September 2007) at 25–26.

algorithm does not remove the possibility that the algorithm discriminates against Māori. The algorithm was based on an assessment of the conviction and sentencing outcomes of tens of thousands of past offenders since the 1980s. The impact of ethnicity goes beyond sentencing and conviction data: each stage of the criminal justice system contains racist assumptions that harm Māori.[87] These assumptions influence whether police apprehend an alleged offender, whether they decide to arrest and prosecute, whether the court convicts that person and what sentence a judge applies. Removing ethnicity from the equation is a step in the right direction but does not preclude the possibility that data is biased.

AI could also amplify the effects of biased data. This problem is commonly seen in policing. If police monitor particular neighbourhoods more closely than others, it is likely these areas will register higher levels of crime, and this may, in turn, result in a stronger police presence.[88] This process creates a feedback loop in policing based on crime statistics. Similar situations have also arisen around the use of AI and technology in local government. In 2011, Boston introduced an app called Street Bump. The app collected data from citizens, who could report issues with road conditions.[89] The idea was to provide the government with real-time information to fix roading problems more efficiently. The result showed significantly more problems in the roads of Boston's wealthiest areas, which led to the targeting of resources in those areas. Conversely, fewer people in poorer neighbourhoods had smartphones. This led to an under-representation of poor neighbourhoods in the data — even if there were more roading issues in those neighbourhoods. These examples illustrate how data collection can inadvertently produce biased data sets, which might, in turn, lead to the production of even more biased data sets.

Finally, algorithms may be biased by design. Algorithms can never be truly value-neutral as they are designed by humans, who are inevitably influenced by their own values and interests. The values algorithms embody will reflect cultural or other assumptions of the software engineers who design them, and these will be embedded within the structure of the AI. For example, if an algorithm were used to identify a person's credit risk, it might include a person's place of birth, where they went to school, where they live and their employment status as factors. The selection of these factors requires a value judgment. The answers to these questions are relevant in assessing whether the government should offer assistance and, if so, on what

87 At 11–12.

88 Rashida Richardson, Jason M Schultz and Kate Crawford "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice" (2019) 94 NYU L Rev 192.

89 Kate Crawford "The Hidden Biases in Big Data" *Harvard Business Review* (online ed, Massachusetts, 1 April 2013).

terms. This problem is not helped by the demographics of AI professionals, as the vast majority are Caucasian males.[90]

2 The Difficulty in Challenging Algorithmic Bias

Under existing law, it is difficult to challenge algorithmic decisions on the basis that they are biased or discriminatory. The first potential route under existing law is the rule against bias. The right to an unbiased and impartial adjudicator is protected under s 27 of the New Zealand Bill of Rights Act 1990 (NZBORA), as well as under the common law.[91] The right aims to ensure impartiality, preventing any person from being a judge in his or her own cause. It generally applies to situations where a decision-maker has some sort of stake or personal financial interest in the decision being made, or a relationship with the parties that means the decision-maker is not impartial.[92] The right can also apply where a decision-maker is predisposed towards a particular result.[93] In the case of algorithmic decision-making, the AI is immune from many of the biases normally considered under this rule. The algorithm does not have a pecuniary interest in the outcome of a case, an association with one of the parties or a strong opinion on a particular area of law or the set of criteria it is applying. However, if the underlying data an AI uses to make a decision is biased, or if the AI computes and analyses data in a biased way, it might be predisposed towards a particular result. Such predisposition jeopardises an individual's right to natural justice.

This position is, however, theoretical. Any reasons provided for individual decisions, assuming AI can provide sufficient reasons, may not actually be useful in proving algorithmic bias. That is because the law on bias and what amounts to sufficient reasons is individualistic. A successful claim on these grounds may provide a claimant with the data about them that was used, the factors that were considered and why, in balancing those factors, the decision was reached. As the court held in *Re Vixen Digital Ltd*, the reasons provided must allow a claimant to "understand the basis for decisions as to be better informed in predicting that which is or is not within the law".[94] However, identifying algorithmic bias goes beyond how an algorithm treats an individual. A problem with algorithmic bias is how AI treats groups, and how it establishes patterns of bias for or against particular groups of people. As an analogy, proof of a single "stop and frisk" incident in the United States would not reveal the greater racial discrimination in the New York policy, where over 80 per cent of those stopped were black or

90 World Economic Forum *The Global Gender Gap Report 2018* (17 December 2018) at 29–32.
91 New Zealand Bill of Rights Act 1990 [NZBORA], s 27; and Smith, above n 31, at 595.
92 Smith, above n 31, at 595.
93 *Loveridge v Eltham County Council* (1985) 5 NZAR 257 (HC) at 264.
94 *Re Vixen Digital Ltd* [2003] NZAR 418 (HC) at [43].

Latino men.[95] Identifying algorithmic bias based on reasons that apply to individuals may be fruitless.

Further, the test for apparent bias is ill-suited to algorithmic decision-making. Apparent bias involves a two-stage inquiry: there must be actual circumstances that have a direct bearing on the decision-maker's bias, and a fair-minded lay observer must reasonably believe these circumstances mean the decision-maker may not bring an impartial mind to the resolution of the case.[96] These questions do not apply to algorithmic decision-making. AI does not have any sort of mind, let alone an impartial one. AI does not have an interest in the outcome of any of its decisions. It is not a question of incentives or motivations — which are at the heart of the test for bias — but instead whether AI has been programmed to have predispositions for, or against, certain parties. It will also be challenging to ascertain the circumstances that might lead AI to be a biased adjudicator. The circumstances that might make an AI decision-maker biased lie in data and code. To prove an AI was predisposed against a particular claimant, an individual would need to show that the data or code used by the AI was biased. In practice, this will be a highly technical question that even AI developers may not be able to answer, let alone the affected individuals.

Moreover, perhaps a "fair-minded lay observer" is the wrong standard for assessing algorithmic bias, given that a fair-minded lay observer may not have the requisite technical literacy. A better standard, in line with the proposal later in this article, is that of the "fair-minded expert observer".

Finally, to the extent it is possible for individuals to identify bias, time will also be an issue. Limitation periods will be prohibitive, given the complexity of the assessment required. Immigration appeals, for example, require a claimant to file judicial review proceedings in the High Court within 28 days from the time the claimant is notified of the Tribunal's decision.[97] These considerations paint a picture of the difficulty of fitting algorithmic bias within the existing legal tests for bias.

Alternatively, decisions could be challenged on the basis that they are discriminatory. The NZBORA gives everyone the "right to freedom from discrimination" and defines discrimination in terms of the grounds provided for in the Human Rights Act 1993.[98] This right would cover discriminatory algorithmic decisions in principle and is a ground for judicial review. However, challenging these decisions would be difficult. One reason is that it is difficult to distinguish between statistical correlations and discrimination. AI breaks data into groups and finds correlations. Characteristics such as race, age or gender may be found by AI to

95 Dillon Reisman and others *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability* (AI Now Institute, April 2018) at 19.
96 *Muir v Commissioner of Inland Revenue* [2007] NZCA 334, [2007] 3 NZLR 495 at [62].
97 Immigration Act 2009, s 247.
98 NZBORA, s 19.

statistically correlate with relevant outputs, such as the risk of recidivism or unemployment. Statistical correlations may reflect past practices of discrimination such as discriminatory policing practices, but they are not necessarily discriminatory in and of themselves.

The correlation–discrimination dichotomy was one of the problems considered in *State v Loomis*, the first case in the United States to consider issues with the transparency of algorithmic decision-making tools.[99] The Department of Corrections had produced a pre-sentencing report for Mr Loomis, who had been charged with five criminal offences related to a drive-by shooting. Corrections used COMPAS, a predictive risk assessment algorithm employed widely across the United States, to help courts determine the risk of recidivism and pre-trial offending. Mr Loomis argued that the lack of transparency around the COMPAS algorithm violated his due process rights. The court rejected his appeal, largely because intellectual property laws prevented disclosure of information on how COMPAS worked. Although a minor point in the litigation, Bradley J held that the use of gender as a factor in the risk assessment served the non-discriminatory purpose of promoting accuracy, and was not itself discriminatory.[100]

A third possible route for a claimant wishing to judicially review an algorithmic decision is showing that a decision-maker took an irrelevant factor into consideration. If irrelevant considerations influence a decision-maker, the exercise of discretionary power may be invalid.[101] For example, if AI took into account generalised data about ethnicity in approving or declining a welfare application, there is scope to argue this would be an irrelevant consideration. Ascertaining which factors are irrelevant, however, would be difficult. First, individuals need to be told what factors were considered. Secondly, decision-makers are entitled to consider irrelevant matters so long as these matters do not materially influence the outcome of a decision.[102] This latter consideration raises the question of weight, which is typically a matter for the decision-maker.[103] Judicial review generally steers away from assessing the substance of a decision. However, it may be impossible to avoid this inquiry when ascertaining how an AI takes the vast number of potential inputs into account. But as discussed earlier, where AI is a black box, it may be impossible to determine how different factors are treated. In addition, Marion Oswald highlights the tension between computer scientists, who generally wish to feed as much data as possible to algorithms to improve their accuracy, and lawyers, who may restrict the factors taken into account to prevent irrelevancy.[104] There may be a trade-off between the

99 *State v Loomis* 881 NW 2d 749 (Wis 2016).
100 At [83].
101 PA Joseph and J McHerron *Laws of New Zealand* Administrative Law (online ed) at [33] and [35].
102 At [35].
103 At [39].
104 Marion Oswald "Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power" (2018) 376 Phil Trans R Soc A 1 at 10.

accuracy of algorithmic tools and the extent to which potential irrelevant factors are taken into account. A rights-focussed approach might wish to exclude those factors, but this might reduce the quality of the decisions made. This trade-off poses an additional layer of complexity in determining whether algorithms are biased or discriminatory.

AI may produce biased decisions in complex but invisible ways. The challenge is, therefore, not just that AI is a black box, but that the black box may mask discriminatory decisions. Under the existing legal framework, identifying instances of bias or discrimination will be difficult. Requiring reasons is not a panacea. However, it would significantly assist individuals and courts in evaluating the fairness and legitimacy of algorithmic decision-making and increase the accountability of these tools.

Keeping a Human in the Loop: the Problems of Discretion and Delegation

The *Algorithm Assessment Report* emphasises keeping "humans in the loop" as a way to militate against the risks of AI in decision-making.[105] There are two main ways this could be done. The first is by tasking "simple" decisions to AI, leaving more complex deliberations to humans.[106] This approach has been applied in ACC's new system for approving claims.[107] The system uses two algorithms: one that determines whether a decision is complex enough that it should be subject to manual review, and "one that predicts the likelihood a claim would be approved, based on historical data".[108] Together, the algorithms automate and expedite the processing of around 90 per cent of ACC's claims.[109] These models either automatically accept a claim or refer it to humans to review. A human officer will then process and decline any claims.[110] Similar systems have also been applied by Immigration New Zealand[111] and local government for rates calculations.[112] Having to ask when algorithms should be used suggests algorithmic decision-making may not always be appropriate. The level of automation will vary depending on the nature and impact of a specific decision. For example, where decisions affect people's lives in significant ways, the need for human decision-making is heightened. In this way, the model presents a choice: in "simple cases" we might be happy to trade full transparency for increased efficiency.

105 Stats NZ, above n 2, at 30.
106 Gavaghan and others, above n 45, at 54.
107 Stats NZ, above n 2, at 36.
108 Gavaghan and others, above n 45, at 20; and Accident Compensation Corporation *Statistical models to improve ACC claims approval and registration process* (21 August 2018) at 7–8.
109 Accident Compensation Corporation, above n 108, at 4.
110 At 4.
111 Stats NZ, above n 2, at 37.
112 See *Northland Regional Council v Rogan* [2018] NZCA 63, [2018] NZAR 507.

However, algorithms might still make mistakes in "easy" cases, and without reasons these mistakes might go unnoticed for years. This occurred in Canada's Ontario Works programme, which used an algorithm to assist frontline welfare caseworkers.[113] The software could generate decisions based on the data entered by caseworkers, ostensibly freeing up time for caseworkers to meet with more complex clients. The software required caseworkers to answer questions about applicants using a drop-down box with a limited set of options. Based on these answers, it automatically decided "simple" welfare cases. However, the software turned out to impose requirements found nowhere within the empowering legislation. For example, it required that every recipient was enrolled in, or had graduated from, secondary school.[114] Yet, while many potential recipients were declined welfare support on this basis, this discrepancy was not identified until well after the payments were made.[115] The program also made connections between different historical recipients in order to identify how much support each household should receive. However, in some instances the software made sole-support mothers dependent on previous household members, such as former partners or their parents, even where caseworkers had identified that these individuals did not live together.[116] Again, many payments were reduced when they should not have been. These examples show that "easy" cases may not always be that easy, and mistakes can fall through the cracks. Without any clear requirement for explanations, screening out "easy" cases will not address issues with algorithmic decision-making.

A second way in which humans may be "kept in the loop" is where AI is used to assist human decision-making. This is the most common way algorithms have been used in the New Zealand public sector so far.[117] Algorithms may be used to support human decision-makers or may recommend an outcome subject to human approval. This instrumental use of algorithms, however, risks misconstruing the real way humans interact with technology. Keeping a human "in the loop" represents a "form of transparency and personal accountability that is more familiar to the public than automated processes".[118] Humans may be able to moderate any mistakes algorithms make, ensuring the preservation of the compassionate and "common sense" elements of decision-making.

This idea assumes humans will understand and challenge algorithmic recommendations. However, such a challenge is unlikely.

113 Jennifer Raso "Unity in the Eye of the Beholder? Reasons for Decision in Theory and Practice in the Ontario Works Program" (2020) 70 UTLJ 1 at 18–19.
114 At 22.
115 Jennifer Raso "Displacement as Regulation: New Regulatory Technologies and Front-Line Decision-Making in Ontario Works" (2017) 32 Can JL & Society 75 at 84.
116 At 85.
117 Stats NZ, above n 2, at 36–38.
118 At 30.

Decision-makers will likely lack the technological knowledge and time required to evaluate whether an algorithm has produced a fair and correct result.[119] Additionally, algorithms have authority. Bias towards automation means human operators tend to "trust the automated system so much they ignore other sources of information, including their own senses".[120] This bias can be partially explained by how technology can reduce situational awareness and the ability to respond quickly and intuitively to challenges. Drivers in "autopilot" mode, for example, respond more slowly to challenges than drivers in full control.[121] People tend to over-rely on automated decision support systems, even if they do not understand how they work.[122]

Over-reliance on automated decision support systems may lead to improper delegation. The proper exercise of discretionary powers is fundamental to the democratic legitimacy of administrative law. Elected officials give discretionary powers to decision-makers with the proviso that they act within the bounds of their discretion.[123] An element of this rule is that decision-makers must not unlawfully sub-delegate their powers.[124] This rule, however, contemplates situations where a human decision-maker asks for another human's assistance in decision-making. It does not envisage the scenario considered here. Where a human decision-maker leans heavily on AI, is them ticking the final box a real and genuine exercise of discretion, or an impermissible delegation of power?

The Court of Appeal considered this question in relation to an algorithm used by the Northern Regional Council in *Northland Regional Council v Rogan*.[125] The appellants were members of the Mangawhai Ratepayers' and Residents' Association and issued judicial review proceedings to challenge the legality of rates charged by the Kaipara District Council and the Northland Regional Council. Based on the requirements in the Local Government (Rating) Act 2002, the Kaipara District Council used an algorithm to determine the rates paid by residents. Section 53(1) of the Act authorised a local authority to appoint a person or another local authority to "collect the rates they assess". The respondent argued that the section required a person to conduct the assessment.[126] The issue was whether relying on an algorithm amounted to an unlawful delegation of Kaipara District Council's discretion.[127] The Court did not find any issue with the use of the algorithm, because there was no element of discretion or evaluative

119 Zalnieriute, Bennett-Moses and Williams, above n 32, at 442.
120 Kayvan Pazouki and others "Investigation on the impact of human-automation interaction in maritime operations" (2018) 153 Ocean Engineering 297 at 299.
121 Gavaghan and others, above n 45, at 37.
122 Linda J Skitka, Kathleen Mosier and Mark D Burdick "Accountability and automation bias" (2000) 52 Int J Human-Computer Studies 701 at 704.
123 Joseph and McHerron, above n 101, at [41].
124 At [41].
125 *Rogan*, above n 112.
126 At [32]–[33].
127 At [27]–[28].

judgment required in producing the amount payable by the ratepayer. There was only one correct answer, calculated by a mathematical rating formula.[128] As a result, s 53(1) did not preclude the use of algorithms for rates assessments; it would have been an "elliptical" way for Parliament to have limited the use of algorithms.[129]

However, the rule in *Rogan* only covers the most simplistic applications of AI: plain and transparent mathematical formulae. The problem is where humans rely on algorithms that consider complex cases and do more than calculate a right or wrong mathematical answer.[130] When decision-makers rely heavily on AI that makes evaluative, complex recommendations, and there is no clear evidence of meaningful human oversight, there is a strong case that the decision-maker's discretion is unduly fettered. It is unclear what evidence of oversight entails: are decision-makers required to prove they understood how an algorithmic tool reached its recommendation? If so, how could decision-makers prove they turned their mind to this and did not merely "tick a box"? Ultimately, a decision-maker cannot be said to have exercised evaluative judgment when it is the AI who has exercised this judgement.

Carltona Ltd v Commissioner of Works outlined a common law exception that allows ministers to delegate their authority to ministerial advisers.[131] Effectively, the department officials are the "alter egos" of their minister.[132] The idea is that administrative efficiencies and realities justify officials standing in the shoes of their ministers and making routine decisions on their behalf.[133] This principle has been applied widely in New Zealand.[134] A similar rule may be developed concerning algorithmic decision-making. However, it is one thing for ministers to look to civil servants for assistance in routine decision-making, and quite another for decision-makers to rely on algorithms instead. Express statutory authority is preferable, as envisaged by the Court in *Rogan*. And in the absence of such authority, AI-assisted decision-making may be found to unduly fetter administrative decision-making powers. This consideration may be a significant limitation on the applications of AI tools in public sector decision-making.

Finally, both scenarios in which humans are kept in the loop are short-sighted. The first scenario, where AI makes "easy" decisions, assumes humans are better suited to making "complex" decisions. However, algorithmic applications have been found to routinely outperform human

128 At [29]–[30].
129 At [34].
130 At [31].
131 *Carltona Ltd v Commissioners of Works* [1943] 2 All ER 560 (CA) at 563.
132 Smith, above n 31, at 697.
133 *Carltona*, above n 131, at 563.
134 See *Bounty Oil & Gas NL v Attorney-General* [2010] NZAR 120 (HC); and *McInnes v Minister of Transport* HC Wellington CP240/99, 3 July 2000.

experts, both in terms of cost and quality of output.[135] Even if algorithms produce results that are only just as "good" as those produced by humans, they will do so at a fraction of the cost. The public purse is limited, and the economic arguments in favour of increased automation in the public sector are difficult to dispute. Further, the second scenario, in which humans provide a final check on algorithmic decisions, undermines the key benefits of using algorithms in the first place — their pace, accuracy and efficiency. Algorithms can make decisions considerably faster than humans. Reducing bureaucratic tardiness is a significant advantage of AI and would have tangible effects on people's lives. Welfare applications could be processed faster, immigration applicants would know whether they could come to New Zealand much sooner and ballot applications to schools could be resolved in minutes.

Retaining human control might consequently mean one of two things. First, the human role in overseeing these algorithms will become tokenistic. A bureaucrat may just tick off a raft of decisions, relying heavily on the AI's recommendation and having no real impact or oversight. Alternatively, if humans *do* have a more meaningful oversight role, the efficiency gains associated with using algorithms will be undermined. Keeping a "human in the loop" would just be an illusory reassurance.

IV A DUTY TO GIVE REASONS

This article has so far identified three risks of decision-making by AI. First, AI is an unintelligible black box. Secondly, there is a risk AI will produce discriminatory or biased decisions. Finally, keeping "humans in the loop" is unrealistic and, particularly without a requirement for reasons, risks unduly fettering decision-makers' discretion. All three risks weaken the ability of administrative law to be an effective check on algorithmic decision-making. The central argument in this article is that the development of a duty to give reasons is a necessary response to the use of algorithms in the public sector. This final section provides a high-level overview of what a duty to give reasons might look like, and why it is the right response to the challenges highlighted in this article. An important caveat is that the development of this duty should not be viewed as a silver bullet — it is not the only solution to the problems that may arise from AI in administrative decision-making. It is, however, an important first step and may be a prerequisite for more advanced and comprehensive solutions.

135 Ford, above n 9, at 134.

Envisioning a Duty to Give Reasons

As a starting point, any duty to give reasons ought to be set out in statute. Legislation would provide certainty to AI developers and claimants, requiring any decisions made or assisted by AI to be accompanied with reasons. However, in addition to an individual-focused right to reasons, a watchdog agency should be created to oversee and adopt a wide-angle view of AI use in public sector decision-making. This suggestion aligns with the recent proposal made in the NZLF Report and has been mooted, but not applied, in other jurisdictions.[136] Armed with technical expertise, a watchdog agency would work alongside government agencies that use predictive algorithms and conduct ongoing monitoring on the use of these tools. The agency could have the power to require the reassessment and redeployment of algorithms where problems are found to arise. For public accountability, the agency's recommendations and findings could also be subject to the OIA regime.

There are several advantages to this approach. Unlike citizens who face significant time, resource and knowledge constraints, the agency would have the time and technical expertise to assess whether AI tools are working appropriately. The agency would address the concern of Zerilli and others that the type of explanation required to address the risks of AI fully would be "too detailed, lengthy, or technical to satisfy the requirements of practical reasoning" by placing the burden of reviewing the algorithms' functioning on experts.[137] Further, a watchdog agency would be able to address the concern that many of the problems with algorithmic decision-making are not identifiable from individual cases. Determining whether algorithms have been biased against a particular group, for example, requires one to identify a pattern of biased discrimination. Individuals are not privy to this information.[138] Even if they are, this information would not be helpful to most claimants who are "mostly too time-poor, resource-poor, and lacking in the necessary expertise" to make use of it.[139] There is also a risk that by releasing too much information about an algorithm to members of the public, the algorithm would be open to hacking or manipulation by users, or copying by competitor companies. This risk is mitigated by a watchdog agency, who would be subject to tight privacy controls. An agency taking a wide-angle approach to algorithmic decision-making would provide effective oversight over these tools.

136 Gavaghan and others, above n 45, at 62–70.
137 Zerilli and others, above n 75, at 668.
138 Privacy Act 1993, s 6(1) IPPs 10 and 11.
139 Edwards and Veale, above n 81, at 67.

Why a Duty to Give Reasons is Necessary

1 An Incentive to Develop Transparent AI

A duty to provide reasons provides a much-needed incentive for developers to create explainable AI. Without any legal obligation to produce explainable AI, and without the knowledge that any AI developed will be subject to agency oversight, there are few such incentives. Although some developers are developing "transparent box" AI that explains in simple terms how decisions are reached, this is a costly and time-intensive exercise. Currently, the incentives are profit-driven. Companies might be able to claim a competitive edge if they pitch transparent AI systems when competing for government contracts. However, it is unclear to what extent the market (including private markets, which drive most development of AI) will, in the medium to long term, require explainability from AI. If there is little organic market demand for explainable AI, then a duty to give reasons is vital. Explainability is unlikely to otherwise become a priority for AI developers. On the other hand, if, in the future, explainability does become a private market concern, there is a risk that the quality and extent of explainability are set by market forces. In other words, non-public actors will determine what "explainability" means in practice. A duty to give reasons, therefore, not only incentivises the development of explainable AI, but also avoids private actors from holding the power to specify what explainability comprises.

The obvious issue with this solution is that developers themselves may be unaware of how AI produces results. The opacity of algorithms is not necessarily a design choice. Rather, AI's opacity is often a product of its complexity. Zerilli and others argue that moving towards more transparent or explainable AI will require a trade-off between the capabilities and transparency of AI systems.[140] Further, the authors consider that requiring high levels of transparency might have a chilling effect on the development of AI, "[preventing] deep learning and other potentially novel AI techniques from being implemented in just those domains which could be revolutionised by them and have the most to gain".[141]

There are indications, however, that transparent AI is possible.[142] To the extent that transparency reduces the quality of AI tools, this trade-off is worth making given the importance of explainable AI. Further, even if AI developers do not have a complete understanding of exactly how their AI works, they are in the strongest position to take steps towards developing more transparent AI now. As AI is not currently as complex as it soon will

140 Zerilli and others, above n 75, at 668.
141 At 668.
142 David Gunning and David W Aha "DARPA's Explainable Artificial Intelligence Program" (2019) 40(2) AI Magazine 44 at 44–45.

be, requiring explainable AI at this relatively early stage is pivotal. Finally, even if the explanation is not perfect, any explanation is better than no explanation at all. Given the difficulties that surround building explainable AI systems, a duty to give reasons would provide the strongest incentive to bring about this development.

2 Democratic Legitimacy of Administrative Law

Finally, the role of reasons in the democratic legitimacy of administrative law takes on heightened importance in the age of algorithms. In the context of human decision-makers, Mashaw has argued that reason-giving is a remedy against the apparent "democratic deficit" of administrative law.[143] One of the reasons is that those who make administrative decisions are typically unelected.[144] Powerful administrative agencies and bureaucrats carry out the policies and legislation set by democratically-elected officers, and judicial review is one of the only checks on the exercise of their powers. Because individuals need to understand decisions made about them to be able to challenge them, reason-giving is a fundamental element of a participatory administrative state that prevents the arbitrary or otherwise improper use of public power. Reason-giving reinforces in people a sense of trust in the administrative state.

The need for reasons is even more pronounced in the context of algorithmic decision-making. People will be less likely to trust algorithmic decision-making when no explanation is provided for decisions. In general, people are wary of AI. People are concerned about the rise of big data and what that means for their privacy; they are wary of big tech companies and of technologies they do not understand.[145] Algorithmic decision-making ignites all these concerns. Public sector decision-makers will feed data about individuals to algorithms. And these algorithms — which may have been developed by big tech companies — might make decisions without reasons.[146] Just the fact of being turned into data points can feel disconcertingly impersonal, let alone not having an opportunity to understand how the decisions were made. This may end up eroding people's sense of dignity and trust in AI.

Further, any trust people have in algorithms already stands on shaky foundations. It is easy to imagine one "stray" algorithmic decision having a

143 Jerry L Mashaw *Reasoned Administration and Democratic Legitimacy: How Administrative Law Supports Democratic Government* (Cambridge University Press, Cambridge (Mass), 2018) at 164 and 183.

144 At 164 and 170.

145 Alison Kay "Why trust in tech giants is eroding, and how it can be rebuilt" (29 March 2018) EY <www.ey.com>.

146 See New Zealand Human Rights Commission *Privacy, Data and Technology: Human Rights Challenges in the Digital Age* (May 2018) at 20–21 for an example of the use of big data by the government.

ripple effect on the wider trust people have in AI. Consider how people's trust in self-driving cars tumbled when the public learned of an accident involving one car.[147] Further, administrative decision-making is one of the most common ways in which people interact with the government. If people do not trust the decisions public sector agencies are making, and have no real means to challenge those decisions, their sense of trust in the government will inevitably be compromised. That, in turn, impacts the democratic legitimacy of the administrative state. A duty to provide reasons provides a bulwark against this risk.

However, there is a case that algorithms might be more democratically legitimate than human actors in terms of the results they produce. Human decision-makers are imperfect. Jennifer Raso's study highlights how frontline caseworkers consistently strayed from the black letter welfare criteria in order to fit some applicants within the rules.[148] In doing so, the caseworkers departed from Parliament's intention in creating certain conditions that applicants needed to meet. Algorithms are much less likely to do this. Algorithms, after all, apply rigid and formulaic mathematical tests to applicants. Although there are myriad ways in which algorithms may display bias, they are far less likely to manipulate particular applications on a case-by-case basis to circumvent legislative rules. In this sense, algorithms are very positivistic, not allowing questions of morality to infuse their decision-making. Algorithmic decisions may more closely follow legislative intent than humans.

Nonetheless, technical compliance might not improve the way people feel about the legitimacy of administrative decision-making. Ronald Dworkin defines decision-making in terms of the exercise of both moral and practical reason.[149] From this perspective, morality is fundamental to the legitimacy of decision-making. In some cases, such as the calculation of rates payments, morality does not play a role. However, in others, such as welfare applications, Parliament appears to be content to give discretion to welfare caseworkers knowing they may make a decision based partly on their idea of right and wrong. This allocation of discretion raises a related normative concern: can complex administrative decisions be reduced to mathematical equations? Algorithms require decisions to be made based on quantitative assessments of value judgments that otherwise may be qualitative. This criticism has already been levelled at the social investment approach to policy, which places a heavy focus on the quantification of benefits and harms of potential interventions. Examples include the quantification of the value of human life in terms of GDP, as well as other

147 Meredith Broussard "Self-Driving Cars Still Don't Know How to See" *The Atlantic* (online ed, New York, 20 March 2018).

148 Raso, above n 115, at 83.

149 Ronald Dworkin "Philosophy, Morality, and Law—Observations Prompted by Professor Fuller's Novel Claim" (1965) 113 U Pa L Rev 668 at 672.

questionable values, such as the monetary value of biodiversity.[150] In the context of algorithmic decision-making, can the value an immigrant might provide to New Zealand be measured solely by GDP? Is it fair to evaluate one person's risk of abusing his or her child based on the data of how other "similar" people treated their children?

These are issues that reason-giving might not ameliorate: people may feel like the "human touch" of decisions is lost. Trust in algorithmic tools will take time to develop but will probably eventuate. Machine learning algorithms will likely continue improving at a rapid pace and become increasingly normalised in our lives.[151] However, central to any such development needs to be a requirement for rationality in algorithmic tools. Reason-giving has long been considered fundamental to the democratic legitimacy of administrative law. This is only heightened with the rise of algorithmic decision-making. People's sense of dignity, of trust in the government and ability to participate in decisions made about them rely on transparent and explainable AI. Reason-giving may raise a host of new issues in the age of algorithms, but is key for the continued democratic legitimacy of administrative law.

V CONCLUSION

This article has highlighted some of the real and important challenges AI presents to administrative law. There is a risk that no one will understand how AI reaches its decisions, that AI will be biased and discriminatory and that the use of AI might unduly fetter a decision-maker's discretion. Each of these risks engages fundamental principles of administrative law. Understanding how AI reaches decisions is a vital element of natural justice, reflecting the importance of human dignity and the ability of individuals to challenge decisions. The rights to an unbiased adjudicator and to be free from discrimination are similarly fundamental. However, the protection of these rights against AI decision-makers requires a level of understanding and proof a layperson is unlikely to possess. There is also a risk that discretion is improperly delegated to algorithmic tools, in a way that is at odds with the rule against sub-delegation.

Values of transparency and openness are fundamental tenets of good governance. However, decision-makers under the current law are not subject to any common law duty to give reasons. While the OIA provides a backstop that allows any person to request reasons be provided, this alone is insufficient to protect against the risks of AI. The increasing use of AI in

150 Thomas Coughlan "Is a life worth $4.7 million?" (21 February 2019) Newsroom <www.newsroom.co.nz>.

151 Ford, above n 9.

decision-making necessitates the development of a duty to give reasons. Therefore, this article has proposed a duty to give reasons that takes effect in two stages. First, a rule that any decision made by an algorithm needs to be accompanied by reasons. Secondly, this article suggests the establishment of a watchdog agency that provides oversight and more technical analysis of algorithmic decision-making tools. This solution is proactive and balances the risks and advantages of algorithmic decision-making in the public sector.

The application of AI to almost every aspect of our lives is inevitable; we are in the midst of a technological revolution. Administrative law cannot stand still in the face of the important challenges and opportunities presented by AI. The rise of algorithmic decision-making in the public sector, therefore, makes the development of a general duty to give reasons essential.