Amelia Retter

Man vs Machine: Accountability Mechanisms and New Technology

Submitted for the LLB (Honours) Degree

LAWS522

Faculty of Law

Victoria University of Wellington

2018

Abstract

Technology is becoming more complex and is increasingly being used in law. Tools to assist in decision making and becoming more complex. It is important to ensure accountability structures keep up with their development so we do not lose control of decision making processes. This paper identifies four types of decision making using algorithms: human decisions, decisions using non-machine learning algorithms, decisions using machine learning algorithms, and decisions where machine learning *makes* the decision. Issues are identified in applying accountability mechanisms for each, focusing on challenges in pinpointing an actor to hold accountable and forums equipped to ask questions. The use of machine learning is a significant hurdle in being able to choose an actor because these kinds of algorithms are opaque and require significant expertise to comprehend. Users do not necessarily know how the machine works and so cannot provide adequate account for their use. Programmers may have to shoulder some of the accountability burden, however they too may be unable to provide complete answers. Likewise, forums may lack knowledge to ask meaningful questions and lack of transparency on the part of the algorithm. Problems identifying these parties in an accountability context need to be resolved for the future as machine learning algorithms become more common.

CONTENTS

WORD COUNT 13,352 excluding cover page, table of contents and bibliography

I Introduction

Advances in technology are occurring at a rapid pace with new technology being used in all areas of our lives. The legal sphere is no exception, and New Zealand has seen increasing use of technology in legal processes, especially algorithms. Public law is needed to ensure that there is accountability for these developments when used in decision making and exercises of public power. In this paper, I will identify issues created by new risk assessment technology when existing models of accountability are applied. I find that especially when machine learning is utilized, there are challenges in identifying the actor because the user of the technology may not be the party best able to respond to interrogation by the forum due to lack of knowledge about how the algorithm works. A pattern also emerges when looking at the adequacy of forums. Forums commonly used to ensure accountability for risk assessment decisions may lack expertise, and may not be able to get all the pertinent information due to the complexity of machine learning algorithms.

Algorithms have impacts on a wide variety of decisions and processes that use public power. A specific area in which algorithmic decision making and maintaining accountability is of concern is criminal justice. Use of algorithms can lead to more accurate assessments and a safer society as a result. Algorithmic technology and machine learning may make it easier to make risk assessments to increase accuracy in bail and parole decisions, and can also impact the length and type of punishment at sentencing. In fact, New Zealand already uses algorithms such as the Risk of Reconviction, Risk of Re-Imprisonment (ROC*ROI) to aid risk assessment decisions.[1] However, when these decisions go wrong there are potentially serious ramifications for offenders and public, necessitating robust accountability processes to reduce the chance of errors being made and to enable meaningful discussion when things do go awry. The criminal justice context in particular necessitates a strong accountability structure because there are serious consequences that flow from these decisions.

The potential dangers of using technology such as algorithms and machine learning without clear parameters for accountability was highlighted recently. Immigration New Zealand faced questions over their use of electronic decision making tools to determine who to pursue for deportation based on risk.[2] Eventually, it was determined the tool being used was a spreadsheet formula rather than a complicated algorithm.[3] Existing accountability mechanisms

[1] Department of Corrections Policy, Strategy and Research Group "Over-representation of Māori in the criminal justice system: An exploratory report" (Department of Corrections, September 2007) at [2.5]; Lyn Provost "Department of Corrections: Managing off enders to reduce reoffending" (Office of the Auditor General, December 2013) at [3.8].

[2] Gill Bonnett "Immigration New Zealand using data system to predict likely trouble makers" (5 April 2018) Radio New Zealand <www.radionz.co.nz>.

[3] Radio New Zealand "Immigration Minister puts controversial profiling programme on hold" (9 April 2018) Radio New Zealand <www.radionz.co.nz>.

were sufficient to address this incident, however, although this instance was able to be resolved relatively easily, it does illustrate the potential for difficulties when the programme is a more complex algorithm.

It is imperative appropriate accountability mechanisms are in place to ensure decision makers are held accountable for decisions made in reliance or partial reliance on technology. Doing so may involve a re-thinking of conventional understandings and processes of accountability because technology such as algorithms can raise questions about who is making the decision and therefore who must be held accountable. Technology can also make it more difficult to use common forums to demand accountability because technology can be difficult to understand and not transparent.

New Zealand does need to be aware of issues around applying conventional accountability mechanisms to new technology, and begin thinking about how to address them, as examples like ROC*ROI and Immigration New Zealand illustrate algorithms are already part of our decision making processes. It is likely they will continue to grow in number and complexity, necessitating effective accountability structures to prevent errors. Part of such an effective structure is being able to identify the actor to hold accountable and an appropriate forum that can pose questions and demand answers.

The first part of the paper will explain what risk assessment is, how it is currently undertaken, and why we may want to use technology to augment this assessment. It will also cover how risk assessment tools are currently used in New Zealand to illustrate their growing salience and why there is a need to implement effective accountability structures. Accountability will then be explained, along with what is required of an effective actor and forum. Finally, a typology of algorithms will be articulated and issues with applying accountability to these types of algorithmic decision making processes will be identified. These relate to the potential inadequacy of using current conceptions of the actor and forum to hold decision makers accountable. The typology will distinguish between individual decision making, basic algorithms, and machine learning algorithms. These distinctions are important because they affect both actors' and forums' levels of knowledge and understanding of the decision making process and therefore their ability to render and demand an account.

While some potential avenues for solving these problems may be mentioned, I do not attempt to solve problems identified or assess any solution's efficacy in this paper. Rather, I aim to define the issue to illustrate how the law and traditional understandings of accountability for the use of public power in risk assessment decisions needs to develop in line with new technologies.

II The Role of Algorithms in Risk Assessment

Risk assessment can relate to many decisions made in the public sphere. A particular area in which they can assist in making these decisions is criminal justice. In the criminal justice context, the role of risk assessment is to make better decisions about safety so dangerous individuals are not released into the community.[4] Another purpose may be to ensure low risk individuals are insulated from harms that stem from incarceration by allowing these individuals to reintegrate. Further, accurately assessing risk will mean potentially scarce criminal justice resources are used effectively on those who are the highest risk,[5] while resources are not unnecessarily spent on those who can safely live in the community despite committing a crime. Those given low risk scores would be more likely to be bailed or paroled and to receive lighter sentences and vice versa for high risk offenders.[6] Risk assessment serves an important purpose by making these, and other, functions easier. It is therefore important to ensure that when it is not operating correctly for whatever reason, answers can be demanded through accountability mechanisms.

Risk has always been a part of criminal justice, whether it is determining whether to release someone on bail, grant parole, or give a community rather than a custodial sentence.[7] Legislation even recognizes the important role of risk assessment. For example, a relevant factor in sentencing is conviction history,[8] which is a means of assessing risk of reoffending indirectly.[9] Assessing this risk is not a new concept;[10] in the early 20th century, a newly-created 21 factor formula proved more accurate at predicting recidivism that psychiatrists given the same task.[11] What is new is that increasingly, criminal justice is moving from solely professional assessments of risk to assessments relying on predictive tools and technology.[12] However predictive tools are often overridden by individuals' judgment of the risk level, because risk assessment tools do not completely replace individual conclusions.[13] Machine learning is the next step in how we make that assessment. We need to ensure the ways we

---

[4] Anna Chalton "Rape Myths and Invisible Crime: The Use of Actuarial Tools to Predict Sexual Recidivism" (2014) 5 PILJNZ 112 at 133.

[5] Waitangi Tribunal *The Offender Assessment Policies Report* (Wai 1024, 2005) at [3.5.5].

[6] Alyssa M Carlson "The Need for Transparency in the Age of Predictive Sentencing Algorithms" (2017) 103 *Iowa Law Review* 303 at 309.

[7] Leon Bakker, David Riley, James O'Malley "Risk of Reconviction: Statistical Models Predicting Four Types of Re-Offending" (Department of Corrections, 1999) at 7.

[8] Sentencing Act 2002, s 9(1)(j).

[9] Chalton, above n 4, at 132.

[10] Melissa Hamilton "Risk-Needs Assessment: Constitutional and Ethical Challenges" (2014) 52 *American Criminal Law Review* 231 at 236; Bakker, Riley, O'Malley, above n 7, at 3.

[11] Robert Brauneis and Ellen P Goodman "Algorithmic Transparency for the Smart City" (2018) 20 *Yale LJ & Tech* 103 at 112

[12] Hazel Kemshall *Understanding Risk in Criminal Justice* (Maidenhead, Open University Press, 2005) at 27.

[13] At 23.

monitor decisions keep pace with how the assessments are made. Thus, accountability should keep pace with new developments such as the use of machine learning algorithms.

To understand why risk assessment algorithms are becoming more prominent and therefore understand why we need to be ensuring that public law accountability can effectively hold them accountable, we need to know why individual assessments alone may be insufficient. There are a number of challenges associated with relying on individuals to make decisions about risk. First, risk assessments may be subject to biases of decision makers. Forecasting risk should ideally be transparent. However, even when giving reasons for a decision reasoning may be intuitive,[14] and there may be underlying biases influencing these intuitions.[15] In this way, decisions may be influenced by extraneous factors without the decision maker themselves even knowing. There may also be inconsistency between judgments. Risk assessment can vary from person to person based on their ability as well as what underlying biases they have.[16] There can therefore be significant differences between assessments of the same risk based on different features of the decision maker themselves.[17] Not only might this reduce public safety if lenient decision makers allow high risk people back into the community, but it may also harm offenders by keeping them in custody unnecessarily if the decision maker is risk adverse and this inconsistency is unfair on those subject to variable determinations. Not only that, risk assessments are often not particularly accurate.[18] These factors can lead to a desire for an objective and statistical assessment as to risk.[19]

Given these challenges with individual decision making, algorithms may be an attractive addition to the process by neutralising some of the concerns associated with individual decision making. A different approach would be to rely more heavily on more complicated technology to predict behavior and make determinations of risk. Such an approach would theoretically remove inconsistency and bias as well as increasing transparency in the sense we would know exactly what information was being considered by the programme. Due to these benefits, it is likely algorithms will become more relevant in New Zealand in making risk assessment decisions.

---

[14] Richard Berk and Jordan Hyatt "Machine Learning Forecasts of Risk to Inform Sentencing Decisions" (2015) 27(4) *Federal Sentencing Reporter* 222 at 222.

[15] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner "Machine Bias" (23 May 2016) ProPublica <www.propublica.org>.

[16] Carlson, above n 6, at 315.

[17] Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey C. Barnes "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality" (2018) 27(2) *Information & Communications Technology Law* 223 at 237; Sam Corbett-Davies, Sharad Goel and Sandra González-Bailón "Even Imperfect Algorithms Can Improve the Criminal Justice System" (20 December 2017) The New York Times <www.nytimes.com>.

[18] Berk and Hyatt, above n 14, at 222.

[19] Carlson, above n 6, at 305.

Algorithms mean a judge, parole board member, or prison employee does not need to assess a defendant's risk of reoffending.[20] Risk assessment tools replace professional assessments by experts with probability calculated by scores based on risk factors present in any given individual.[21] Risk assessment tools typically work by collecting data about (for example) previously paroled individuals and creating statistical likelihoods a person will reoffend and thus their suitability for parole.[22] The same can be applied to questions as to whether someone will offend on bail, or if they should be released rather than receiving a custodial sentence.

New Zealand already currently uses risk assessment tools. The Court in *R v AM* did not question the use of risk assessment tools at sentencing.[23] Algorithms also operate in the bail context. The Parole Board may consider anything relevant to public safety,[24] which includes any risk assessments that have been made in order to inform them about risks to public safety.[25] Currently, New Zealand employs several risk assessment tools. ROC*ROI is used as the Department of Corrections' risk assessment method[26] and the Automated Sexual Recidivism Scale (ASRS) is used to assess the risk of sexual recidivism.[27] Corrections also uses professionals who administer questionnaires designed to measure factors which may lead to recidivism.[28] New Zealand is thus already using risk assessment tools and it is likely these will further develop as technology becomes more advanced.

Algorithms, despite their benefits, have pitfalls which will become apparent when accountability frameworks are applied to them. Therefore, there is a need to ensure we will be able to use these tools while being certain we can effectively hold actors accountable for their use.

III The Role of Accountability in Public Decision Making

To assess how accountability mechanisms might apply to risk assessment algorithms, it is first important to understand what is meant by accountability. Accountability is difficult

[20] Kirsten Martin "Ethical Implications and Accountability of Algorithms" (2018) *Journal of Business Ethics* at 7.
[21] Kemshall, above n 12, at 28
[22] Carlson, above n 6, at 305.
[23] *R v AM* [2010] NZCA 114 at [143].
[24] Parole Act 2002, ss 7(1) and 7(2).
[25] Chalton, above n 4, at 136.
[26] Department of Corrections Policy, Strategy and Research Group, above n 1, at [2.5]; Provost, above n 1, at [3.8].
[27] Chalton, above n 4, at 141.
[28] At 141.

to define,[29] but includes ideas of people holding their government responsible, fairness, and responsivity.[30] Accountability necessitates an external body to which the one accounting for their actions must answer to.[31] In a broader sense, accountability also includes limits on government action and being able to engage the government in discussion with the public to ensure the people continue to be heard between formal elections.[32] However this is still a broad concept.

A more specific definition of accountability is: [33]

> a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgement, and the actor may face consequences.

I will be using this definition because it provides a clear framework for assessing the necessary elements of accountability. Under this conception, accountability has three components: first the actor must be required to give information, explanations and justifications, second the forum must be able to question the actor, and finally the forum passes judgment.[34] Usually judgment will involve a consequence and it is this that moves the giving of information to an accountability relationship.[35] "Consequence" encompasses more than just sanctions, because "sanction" conveys a negative idea but accountability is broader, as bodies like the Ombudsman cannot sanction formally and sometimes actors are praised.[36] Thus accountability is a specific type of relationship between separate parties involving dialogue and a result.

Accountability is important because it is a way to ensure the use of state power is monitored and constrained where appropriate.[37] It legitimizes the government and promotes confidence because if the public can ask questions, they will get explanations that bolster policy.[38] Effective accountability requires information to be available and for people to be able to ask questions and demand answers.[39] Public dialogue is important, whether through the

---

[29] Mark Bovens "Analysing and Assessing Accountability: A Conceptual Framework" (2007) 13(4) European Law Journal 447 at 448
[30] At 449.
[31] Richard Mulgan "Accountability: An ever expanding concept?" (2000) 78(3) *Public Administration* 555 at 555.
[32] At 556.
[33] Bovens, above n 29, at 450
[34] At 451
[35] At 451
[36] At 452
[37] At 462
[38] At 464
[39] Mulgan, above n 31, at 567

courts or the media or the House, because accountability involves dimensions of explaining and justifying actions and for these explanations to be critiqued.[40]

Often we need accountability mechanisms because power has been delegated. Those elected have the mandate to make decisions but because they lack time or expertise to make all decisions, they delegate some power to others.[41] Voters delegate political power to representatives, who in turn delegate to the executive, who likewise pass on some power to officials.[42] Accountability to the ultimate principals (voters) is indirect through the chain of delegation.[43] However with giving away power comes risks that those who are given responsibility will not use their power appropriately and this is more likely when the principal does not have specialist knowledge about how the delegated task works.[44] To ensure those who are delegated to are using their power appropriately accountability mechanisms can be used.[45] It will be difficult to control the exercise of power that has been delegated if there are no means by which people can question and interrogate actions.[46]

Actors who use them to make decisions are held accountable for the final decision, but also how they come to that decision and the tools they use. Taking the above definition, we can look at the decisions actors make currently within the system to see whether users of risk assessment algorithms would be subject to accountability mechanisms. Actors could be those making decisions which utilize risk assessment tools. Regardless, there are potentially other actors who may be better targeted for accountability, due to issues with the complexity of the algorithms – an issue which will be discussed below. These actors in their decision making have an obligation to provide reasons for their conduct, because of the importance of their function and the context in which criminal risk assessment algorithms are used. Forums, for example appeals or judicial review of decisions, are opportunities for questions to be posed and judgement to be passed. The public may also perform this function. Through this process, consequences are imposed - for example, having a decision returned to the individual decision maker, or the decision being overturned on appeal.

Accountability structures may be ill equipped to deal with new technologies. Law may be too slow to catch up to technology,[47] so current accountability mechanisms may be inadequate to ensure accountability. The law of accountability has not kept up in terms of

[40] At 569
[41] Kaare Strøm "Delegation and accountability in parliamentary democracies" (2000) 37 *European Journal of Political Research* 261 at 267
[42] At 284
[43] At 284
[44] At 270
[45] At 267
[46] Mulgan, above n 31, at 566
[47] Ryan Calo "Artificial Intelligence Policy: A Primer and Roadmap" (2014) 51 *UC Davis L Rev* 399, at 428

algorithms, as current accountability mechanisms are tailored towards human decision makers.[48] State frameworks for accountability cannot easily accommodate machine learning and this is a serious challenge going forward.[49] Nevertheless, we should not necessarily try to stop the use of algorithms, as we must recognise the reality that technology is increasingly playing a role in our everyday lives and in the legal profession. Instead, we should think about how existing understandings might be disrupted and how to adapt to accommodate the new technology. We need to be able to have confident users can be held to account if the challenges do eventuate.

IV Four Types of Algorithms

Given that existing accountability structures might be disrupted by the use of algorithms in public decision making involving risk assessment, it is important to distinguish between types of algorithms. Algorithms can vary in their complexity, from a simple formula on a spreadsheet to an intricate machine learning programme with millions of decision points.

There are two types of algorithms that could be used in the criminal justice context. First, checklist style algorithms take several factors and assign them values, resulting in a score which is correlated to risk.[50] The checklist system aims to digitalise an expert's determination by translating rules a human would consider into code to copy existing professional processes.[51] Thus, features linked to crime would be counted consistently where otherwise individuals may follow the same process but ascribe different weightings to different factors.

There are also more complicated algorithms which use machine learning. Machine learning *creates* rules based on the data to predict future outcomes.[52] Predictive algorithms find correlations between outcomes and factors inputted into the algorithm and machine learning tests which of millions of possible correlations best fit the data.[53] The predictive algorithm process can find connections that otherwise would not be discernible.[54] In this way, the algorithm can transcend what could be programmed in and goes beyond how a human would decide, creating its own process to determine which inputs are most relevant to risk and

[48] Joshua A Kroll, Joanna Huey, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson and Harlan Yu "Accountable Algorithms" (2017) 165 *U Pa L Rev* 633 at 636.
[49] Cary Coglianese and David Lehr "Regulating By Robot: Administrative Decision Making in the Machine Learning Era" (2017) 105 *Geo LJ* 1147 at 1153
[50] Megan Stevenson "Assessing Risk Assessment in Action" (2018) 103 *Minnesota Law Review* at 9.
[51] National Science and Technology Council "Preparing for the Future of Artificial Intelligence" (United States Government, October 2012) at 8.
[52] At 8.
[53] Brauneis and Goodman, above n 11, at 113.
[54] At 115.

determining factor weightings accordingly.[55] Machine learning is about the system improving in its ability to recognise patterns and implement new connections between data.[56] The resulting predictions are not able to be foreseen by the programmers and humans would not be able to identify the connections in the data the algorithm discovers.[57] Consequently, the actual decision the algorithm makes is not made by a human because the human has not given the machine the formula which it must use to decide.

Whether existing accountability structures are sufficient to respond to mistakes or concerns about risk assessment depends on how the risk assessment is made. To that end, there are a number of scenarios which have differing degrees of technological input into the decision making process and different levels of human control, which result in different challenges in identifying the actor and appropriate forum.

The first type of risk assessment decision would be a completely human controlled decision. Risk assessments were initially conducted by experts based on their experience as professionals.[58] These kinds of decisions rely on expertise and do not involve any technology assisted assessment of risk. In such a process, humans use their expertise to make connection between variables that they have observed in the past to be relevant. Here, the level of human input is complete and a human has absolute control over the whole process because they make the decision unassisted. A good example of such a decision in a risk assessment context is a judge making a sentencing decision before the advent of risk assessment calculations. They would use their past experiences with offenders to determine the most appropriate sentence. Similarly, law enforcement officers acting in the moment may use their professional experience to determine whether someone should be arrested or whether they can just be given a warning.

Secondly, there is the current status quo of human controlled formulas. The decision making process in this instance involves human written formulas in the checklist style described above. Professionals making assessments evolved into using statistics to make risk assessments. As early as the 18th century, mathematics was being used to calculate the probability of particular outcomes.[59] Subsequently, professional assessments evolved into

[55] Stevenson, above n 50, at 10.
[56] Calo, above n 47, at 405
[57] Carla Swansburg "Artificial intelligence and Machine Learning in Law: The Implications of Lawyers' Professional Responsibilities for Practice Innovation" (2018) 60 *CBLJ* 385 at 386
[58] Hamilton, above n 10, at 236.
[59] Kemshall, above n 12, at 28.

using statistical analysis to generate scores based on factors linked to recidivism. [60] These scoring tools were combined with professional opinions into a hybrid risk assessment.[61]

Human controlled formulas take the traditional decision making process of professionals and augments it by using objective, mathematical processes to turn factors into a risk score. As explained above, these tools just turn an expert's considerations into a universally applicable formula so all individuals are measured using the same weightings.[62] These tools encode factors a human would consider and give consistent weights to variables across individuals, meaning that biases in decision makers are not exacerbated. There is a high level of human input because the process is based on human decision processes and the connection between variables is determined by humans. Furthermore, there is still a significant degree of human control over the decision because the human decision maker has the ability to disregard the risk assessment if other factors that the formula does not take into account suggest the decision should be more lenient or harsh. The formula is one tool of many which the human can use to make the decision. The generated scores inform but do not determine the outcome of the decision.

These sorts of tools are now being used at many stages of the criminal justice process in particular.[63] Current tools such as ROC*ROI operate in this way, where parole, bail and sentencing decisions are made taking into account these risk assessments. Another example is the Immigration New Zealand spreadsheet. Immigration New Zealand used a model which ranked individuals based on data to suggest which groups of over stayers were more likely to commit crime, among other things.[64] This allowed for faster deportation of those determined by the model to be similar to previous individuals who had caused problems, thus enabling faster deportation before they behaved similarly.[65]

The third decision type is one which uses machine learning but a human is the one who makes the ultimate decision. There is less human input because the machine generates its own formula and finds its own correlative factors, rather than copying a human process. Like human controlled formula, some degree of human control can be retained because the risk score generated by machine learning can be accepted or rejected, or given more or less weight by the human decision maker. Despite theoretically retaining control, as will be discussed below, the level of actual human control may be lessened due to factors such as automation bias. Thus there is a degree of control but it may be less significant than it first appears. COMPAS, a

[60] Hamilton, above n 10, at 237.
[61] At 237.
[62] National Science and Technology Council, above n 51, at 8.
[63] Hamilton, above n 10, at 238
[64] Bonnett, above n 2.
[65] Bonnett, above n 2.

machine learning risk assessment tool used in the United States in bail, parole, and sentencing decisions is an example.

Finally, there is autonomous machine learning. Autonomous machine learning uses the same tool as human controlled machine learning. The difference is that when machine learning is autonomous, there is little to no human input because the algorithm makes the final decision. Issues may be created for accountability under this model because there may be no human decision maker to hold accountable.

	Human controlled	*Human controlled formula*	*Human controlled machine learning*	*Autonomous machine learning*
Tool	Human logic.	Formula written by a human.	Machine learning created formula.	Machine learning created formula.
Human input	Complete (human makes decision unassisted).	Partial (human makes decision using a risk assessment score generated by a human created formula as one factor relevant to the decision).	Partial (human makes decision using a risk assessment score generated by a machine created formula, as one factor relevant to the decision).	Limited (the machine learning algorithm makes the decision using its generated formula).
Human control	Human has ultimate control.	Human retains a significant degree of control.	Human retains a degree of control.	Human has no control.
Example	In the moment police assessments of risk	Immigration New Zealand (spreadsheet formula), ROC*ROI	COMPAS, HART	Programmes like COMPAS if they relied on in isolation

V Applying Accountability to the Four Types

We will now move into considering whether accountability structures can be adequately applied to all these types of decision making when there are different levels of algorithm reliance. The efficacy of accountability structures can vary depending on which type of decision making process is adopted. Different conceptions change who is involved in making the decision and who could have the knowledge to ask probing and meaningful questions about the decision. Taking each, we can think about what the actor and forum does and challenges that might exist as we begin to use the more complex and less human controlled methods. Much law around accountability assumes that decisions are made by individuals, rather than technology, which is why machine learning sits uncomfortably with current accountability mechanisms.[66] The following sections will focus more closely on the latter two categories, but particularly the third. The reason for this is the first two categories are already

[66] Coglianese and Lehr, above n 49, at 1153.

in operation and accountability structures can be applied sufficiently. The third category will be the focus because it is the next step in using technology in risk assessment and accordingly the most likely to cause issues in the future. The last category while interesting is currently unlikely to eventuate, as it is not likely that fully autonomous decision making would be entrusted to machines.[67]

For each type of risk assessment scenario, two questions will be asked: what are the challenges with identifying the actor, and what are the challenges with identifying the appropriate forum? First, identifying the appropriate actor is critical because otherwise accountability will be meaningless. Identifying the actor is vital because we may not get accurate answers and sanctioning someone with no meaningful control over the process will mean there may not be any change to processes, making it more likely the same mistake will be made again. Second, identifying forums is important because the forum needs to be able to ask questions that will give rise to meaningful answers and ones which are relevant to the issue. They also need to be able to critically analyse answers given by the actor to assess whether they are sufficient. Finally, the forum also needs to pass judgment and impose a consequence. Using an inappropriate forum may mean that the consequence is not appropriate or that judgments are made inaccurately due to misinformation or misunderstanding.

A Human Controlled

We turn now to the first question of whether there are any significant challenges in identifying the actor when looking at human controlled risk assessments. When the decision is made by a human independently of technology, it is clear who the actor is. The clarity exists because we can pinpoint exactly who made the decision and since they relied on their own knowledge and understanding of how factors impact risk levels, [68] there are no third parties who may have a hand in the outcome. We also know that they understand the process and what they have taken into account, so they are the best person to be accountable for their decision because they can thoroughly answer questions from the forum.

Aside from the actual decision maker, there are other actors that may be effective to render account in these circumstances. Determining who the actor is in algorithmic decision making is a variation on the problem of many hands. There are potentially multiple agents and principals and thus people who can be held accountable along the chain of delegation.[69] These concerns can be sidestepped by holding the wider organisation itself to account rather than particular individuals.[70] We look to hierarchical structures within the organisation to determine

[67] Berk and Hyatt, above n 14, at 226.
[68] Hamilton, above n 10, at 236
[69] Bovens, above n 29, at 457
[70] At 458

who to hold to account. [71] For example, usually the Minister in charge of the department will be the one to be held accountable when something within the department goes wrong.[72]

The particular decision maker need not necessarily be the party targeted for accountability. Instead, the chain of command within the department means that individual decision makers may be subsumed into the organisation as a whole which has a representative such as a Minister or Chief Executive to answer questions on the organisation's behalf.[73] Nevertheless, within this structure, individuals may be asked to render account internally, so there is a duality of actors being both the representative externally and the forum internally.[74] This is usually how actors are identified for public decision making.[75] Again, this is easily applicable when the decision is human controlled. Even if the individual is not the one rendering account, superiors along the chain of delegation can ask the actual decision maker for information internally and then become the actor themselves when facing external forums.

So, there are few challenges in finding the actor in human controlled decisions, but it is equally important to be able to identify an effective forum. The forum here will depend on which actor we are talking about, whether it is the representative, or actual decision maker. In the case of the latter, the forum is likely to be a political superior and the forum operates along the chain of delegated power. [76] Ministers as a forum for accountability operates along the chain of delegation.[77] Those who have been given delegated power are accountable to elected individuals, who in turn must answer to the public who elected them.[78] For example the official using the algorithm to assist the decision would be accountable to the Minister who is accountable to Parliament who is accountable to the public. This accountability is seen as a way of stopping unfettered exercise of power and prevents the agency being drawn into politics, which ultimately is for the Minister to handle by justifying policy implemented by the department.[79] Since these superiors are somehow connected to the actor (for example someone working in the same organisation or a responsible minister), they are likely to have some knowledge of how the decision process works and relevant background information. Consequently, they are capable of asking pertinent questions and understanding the issue to be able to pass judgment and impose consequences.

---

[71] At 458
[72] At 458
[73] At 458
[74] At 458
[75] At 458
[76] Amanda Sinclair "The Chameleon of Accountability: Forms and Discourses" (1995) 20(2/3) *Accounting, Organisations and Society* 219 at 225.
[77] Bovens, above n 29, at 455
[78] Sinclair, above n 76, at 225.
[79] At 225.

When the actor is the representative or Minister, the forum is more likely to be the public or courts. Here, the forums often have formal information about how the decision was made as the decision maker must usually provide reasons that can be scrutinised.[80] While the decision may involve expertise (for example psychological theories), they are likely to be able to be explained clearly to the forum so that they can then perform their function.

Overall, simple risk assessment decisions can be relatively easily subjected to accountability. It becomes more complex when tools are added.

B Human Controlled Formula

Accountability can work for these kinds of tools too. The Immigration New Zealand example explained earlier is a good illustration of how human controlled formula can be successfully subjected to accountability demands. The Immigration example may illustrate that current accountability mechanisms can actually work quite well when dealing with simple risk assessment and decision making tools.

Presently, decision makers do not completely rely on predictive algorithms and their predictions may be ignored or accepted.[81] This is desirable because unlike humans, algorithms can only consider what they have been told to. While this may be a good thing to eliminate discriminatory attitudes or unfair biases, it also means that they cannot include context in their decisions in the way that humans can. Algorithms may undermine unique assessments of individuals and their particular risk based on circumstances that the algorithm may not fully appreciate.[82] There are limits to what algorithms can consider, confined to pre-programmed variables, whereas a human has far more discretion to consider other individual factors.[83] Models do not have all information available to them and so cannot completely replace human decision making.[84] Thus we may want to be able to respond to unique circumstances that are not necessarily reflected in statistics and information about the person that can be inputted into the programme by retaining human discretion.

Due to the degree of control that the human still has over the process, the actor is relatively easily identified. Accountability and the ability to justify the use of the decision

---

[80] Frank Pasquale and Glyn Cashwell "Prediction, Persuasion and the Jurisprudence of Behaviourism" (2018) 68 U Toronto JL 63, at 66.

[81] Brian Sheppard "Warming up to Inscrutability: How Technology Could Challenge Our Concept of Law" (2018) 68 *U Toronto LJ* 36 at 42

[82] Eric Holder, U.S. Attorney General "Speech Presented at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference, Philadelphia, PA" (2015) 27(4) *Federal Sentencing Reporter* 252 at 254.

[83] Jason Tashea "Risk-assessment algorithms challenged in bail, sentencing and parole decisions" (March 2017) ABA Journal <www.abajournal.com>.

[84] Oswald, Grace, Urwin and Barnes, above n 17, at 230.

making tool is augmented by the fact the actor can see clearly how it works. The human decision maker can do this by looking at the weightings given to each variable. Consequently, they know exactly what the algorithm has taken into account and therefore what contextual factors may need to be added. What context needs to be accounted for will also impact how much weight is given to the risk score. The user has control over how much weight to give the risk assessment score in relation to other contextual factors. As a result, they can still be effectively held accountable for the overall decision because they can talk about why they gave the level of weight they did to the risk assessment tool. The decision maker knows that this is only a tool that can be considered in context with other information not explicitly included in the weightings. It is therefore effective to hold the human decision maker accountable for how they use the decision making tool because they both know what was considered and can adjust the weight they give to the score relative to other contextual factors not included in the formula.

Other actors are also available to provide effective account. As in human controlled decisions, others along the chain of delegation may be called upon to be the actor. As with the previous category, these parties can ask the actual decision maker for information to relay to the external forum.

For an illustration of how actors can be effectively found in these kinds of situations, Immigration New Zealand can be analysed. Immigration New Zealand knew how the formula that they used worked as they had created it on Excel. They could see what factors were considered, how they were weighted, and could adjust based on further contextual factors if necessary. Therefore they were able to give answers during the accountability process that ensued. Due to this, they are the appropriate actor to demand answers from.

However, the example also illustrates that actors we usually hold accountable for government action may not actually be aware of what their departments are doing.[85] The Immigration Minister was unaware of this practice, showing how important accountability along the chain of delegation is. This may create some issues in identifying the actor if the department's practices are not known to the responsible minister. However, as was the case here, it is likely to be easy for the Minister to seek explanations from those who answer to them and relay that information back to other forums. Since Immigration New Zealand were the creators of the formula and knew how it worked, they were able to give answers to the Minister to pass on so the forum so could determine how to proceed.

The next question then is whether forums that are appropriately equipped to deal with this information can be identified. These risk assessment tools are often straightforward formulas that we can understand and forums regularly deal with them. Courts review decisions

[85] Tom Pullar-Strecker "'Urgent' algorithm stocktake to show how government uses our data" (24 May 2018) Stuff <www.stuff.co.nz>.

which involve decision making tools,[86] the public can demand answers, and Ministers also can seek clarification and assurance.[87] For example, ROC*ROI and another decision making tool were the subjects of a Waitangi Tribunal Report where formulas potentially using racial data were scrutinised.[88] The statistical formula is known and finite, and can be dissected to see exactly how different factors interact to produce the final risk score. The forum can see this and analyse it. For this reason, common forums like the public, political superiors, and courts can understand them and ask intelligent questions about their use and justification. There are no issues of whether the variables used in the calculation are discoverable and no issues of complexity that make existing forums ill-equipped to deal with these types of algorithms.

Again, a good illustration is the Immigration incident. The existence of the algorithm in this example was only discovered by the Immigration Minister Ian Lees-Galloway when the media alerted him, and he then requested further information from his department.[89] The public and the media demanded answers, as did Mr Lees-Galloway, and were able to get answers. It was determined that the immigration situation did not involve a complicated algorithm or modelling tool, rather just a spreadsheet used to rank people.[90] Once clarification was given on how the algorithm operated, there was no issue of understanding and forums were able to ask follow up questions. As a result of this event, Immigration New Zealand is no longer using the tool and a review of how government departments are using algorithms has been ordered.[91] The outcome highlights that accountability processes may be initiated and driven by the general public and the media, rather than formal procedures. It also shows that such forums can give rise to consequences, in this case a review of the use of algorithms. It is therefore likely that existing forums are adequate for dealing with simple formulas.

The public are also a potential forum, although with less direct means of accountability. Public accountability does not have the same formal chain to the public as political accountability where accountability is channelled through a responsible representative, rather it is directly linked to the public.[92] This might be for example addressing the public through media releases and answering journalists' questions.[93] However, there is now more direct interaction between the government and the public in the sphere of criminal justice as a result of increasing distrust of professionalism and the rise of penal populism.[94] Increasingly, we are

[86] See for example *R v AM*, above n 23.
[87] Bonnett, above n 2.
[88] Waitangi Tribunal, above n 5.
[89] Lincoln Tan "Immigration NZ's data profiling 'illegal' critics say" (5 April 2018) New Zealand Herald <www.nzherald.co.nz>.
[90] Radio New Zealand, above n 3.
[91] Pullar-Strecker, above n 85.
[92] Sinclair, above n 76, at 225.
[93] At 226
[94] Kemshall, above n 12, at 44.

relying on public groups to assist in demanding accountability because of lack of trust in formal government.[95] Public reporting and the availability of information on the internet has facilitated this but it may be difficult to pass judgment and sanction,[96] particularly if they have to wait until the next election, by which time the issue may have fallen off the agenda or faded into the background. Elections may also be a crude tool to express displeasure at specific decisions because they involve a range of issues. Nevertheless, the public may have other ways of ensuring that a consequence is imposed. In the Immigration example, a result has been the government ordering an inquiry into the use of algorithms in the public sector. Thus we can see that there are specific and direct means by which the public can have an effect on actors to increase accountability.

The efficacy of the forums in the Immigration example was however likely to be due to the fact that the public and the Minister were able to understand the tool. Furthermore, the actors were able to be held responsible because they too knew how it worked. New Zealand has already faced a potential accountability issue with algorithms, but the spreadsheet was not the kind of advanced technology that machine learning algorithms represent. Predictive tools are becoming more complex and autonomous in their calculations, potentially signalling a new stage in risk assessment. Unlike the Immigration example, these tools may not be able to fit as clearly into the accountability structures. So, how does the model apply to other, more complex risk assessment tools?

C Human Controlled Machine Learning

If and when we take the next step of utilizing more complex algorithms such as machine learning, we should not see algorithms as a perfect decision maker capable of replacing human discretion.[97] Instead, they should be used in conjunction with traditional discretionary decision making.[98] For example, use of the Harm Assessment Risk Tool (HART) in England operates on the philosophy that machine learning algorithms should be used alongside human officials. The police use HART to set a risk level of offenders, although it is only supposed to be advisory and police retain their discretion.[99] Currently in the United States, at least 10 states are using algorithms to help determine whether an individual should be bailed or paroled.[100] One programme is Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). In *Wisconsin v Loomis*, the Court stressed the importance of caution when using

[95] Bovens, above n 29, at 457
[96] At 457
[97] Oswald, Grace, Urwin and Barnes, above n 17, at 232
[98] Berk and Hyatt, above n 14, at 226.
[99] Oswald, Grace, Urwin and Barnes, above n 17, at 225; BBC "Police warned about using algorithms to decide who's locked up" (16 November 2017) BBC <www.bbc.com>.
[100] Richard Gifford "Legal Technology: Criminal Justice Algorithms: AI in the Courtroom" (2018) 38(1) *The Proctor* 32 at 32.

algorithms in risk assessment decisions.[101] They emphasised that risk scores are not to be the conclusive factor.[102] Remembering this is important, as risk assessment tools are used alongside judgment of the decision maker and are not conclusive. The current consensus appears to be that humans should retain ultimate control of the decision.

As a result of the retaining of human oversight under this model of algorithmic decision making, machine learning tools with human oversight may at first blush appear to be analogous to formulas used in conjunction with human control. However, as machine learning algorithms are more complex and inscrutable, they present significant challenges that change how we identify the actor and forum. This may disrupt how we conceptualise accountability in these cases.

Actors may be more difficult to identify when using machine learning, even when it is human controlled. Whereas in human controlled formula the user can be held accountable for the decision because they know what variables have been considered so can adjust the significance to be given to the score, users may not be able to be held accountable for machine learning risk assessments because they do not know what the machine has deemed significant. As a result, they may not know what contextual factors to add into their decision. Consequently, we may have to cast the net wider to find an actor who can meaningfully respond to questions and give account. There are a number of people who may be responsible for the decision, depending on numerous factors.

One way of dealing with algorithms is to see them as wholly the responsibility of the users, those who have created them, or those who own them. [103] For now, this seems the most likely response.[104] This is consistent with current approaches where liability falls on to those who have control of the object.[105] Nevertheless, there are still challenges in pinpointing exactly who is the best person to pursue for accountability.

As is the case in the previous two conceptions, individuals and political superiors may be used as actors to provide answers when accountability is demanded.

Currently, the actor can be conceptualised as the user of the algorithm, as they retain ultimate control of the decision and can override the algorithm. In the human controlled formula model, this justified holding the user accountable as the actor. However, the degree of

[101] Tashea, above n 83.
[102] *Wisconsin v Loomis* 68 2015AP157-CR (Wis 2016) at [44]
[103] Iria Giuffrida, Fredric Lederer and Nicholas Vermeys "A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and other Technologies will Affect the Law" (2018) 68 *Case W Res* 747 at 763-764
[104] At 769
[105] At 764

this control and discretion may be called into question when using machine learning with the effect of reducing the justification for holding users accountable. There are a number of reasons for saying that the use of machine learning lessens the control of the user over the decision and therefore appropriateness of holding them to account.

Failing to understand how the algorithm works may reduce the quality of answers that any user can give in rendering account, despite the fact the user in theory retains ultimate control. Where algorithms are used to make risk assessment decisions, public servants may no longer have meaningful control over decision making process because they do not understand how it works.[106] Lack of understanding stems from the fact that we do not know what variables the machine has deemed significant or how these interact. Since they do not understand how the algorithm works, the user as an actor will be less able to provide meaningful answers. This is different from human controlled formula situations because the formula in that case is knowable to the user.

The second reason that users may be an ineffective actor is ambiguity. It may become difficult to determine whether the decision has in effect been made by the machine. It can be hard to tell whether the human decision maker has uncritically relied on the algorithm's judgment, or whether the human has deferred to the machine because it is the most sensible course of action as a result of a considered process.[107] If public officials do not know how an algorithm works and what factors the machine learning processes have identified as relevant and significant, the human decision maker does not know what factors within their own decision making processes they will have to account for so have to follow the algorithm's recommendation or just revert to relying on their discretion due to lack of context.[108] In this scenario, it is more difficult to say that the human "decision maker" should be held accountable because they did not make the decision – the machine did.

The third reason that the degree of control over the decision and algorithm may be less significant than it first appears is because of automation bias. Despite the algorithm being only one factor in a decision maker's arsenal, it can be hard to act against the algorithm's recommendation.[109] Automation bias involves the idea that people are more likely to trust a machine's judgment,[110] even when the user thinks there may be an error.[111] We do not tend to look for more information because we trust technologies and their outputs are adopted

[106] Brauneis and Goodman, above n 11, at 126.
[107] Glenn Cohen and Harry S Graver "Cops, Docs, and Code: A Dialogue Between Big Data in Health Care and Predictive Policing" (2017) 51 *UC Davis L Rev* 437, at 455.
[108] Brauneis and Goodman, above n 11, at 127.
[109] BBC above n 99.
[110] Oswald, Grace, Urwin and Barnes, above n 17, at 238
[111] Danielle Keats Citron "Technological Due Process" (2008) 85(6) *Washington University Law Review* 1249 at 1271

uncritically.[112] Using technology to augment human decision making may be a good approach but it will become more difficult to retain personal decision making as algorithms become more complex.[113] Because of the complexity and apparent superiority of the algorithm, decision makers may feel that they have no choice but to accept its risk score, especially as technology continues to advance and become more sophisticated.[114] As a result, a suggestion is turned into a final decision all because the machine is trusted to give the correct assessment.[115] Consequently, decision makers defer to the technology even when their expertise may point in the opposite direction.

On a related note, if officials are constantly relying on machines to come to a conclusion, they are likely to lose critical skills.[116] Thus holding them accountable is difficult as they no longer have the ability to answer complex questions about their decisions. If reduction in expertise results in deferring to the algorithm, it may well support an argument that we should look beyond the user of the algorithm as they do not understand the how the decision they are making works.

The above issues also apply to ministers if accountability is demanded from them as an actor. Often, politicians may have to accept responsibility for actions of others when they do not fully understand the technology.[117] However, if those users of the technology still cannot give clear answers, the chain of delegation does not work in the same way as it does for human controlled formulas. Therefore similar issues make effective accountability difficult whether there is an individual or organisational application of responsibility. Not only does the Minister not have the knowledge themselves, due to lack of knowledge on the part of the operators they may also not be able to obtain answers from staff.

Another reason users may have difficulty being effective actors relates to consistency. Algorithms are potentially useful tools because inconsistencies can be evened out between decision makers to reduce arbitrariness.[118] The same inputs are relevant and considered for every person in the same way so that individual perceptions and inconsistencies in risk acceptance are no longer influential on the outcome, at least until an individual may override the recommendation of the algorithm. But this could suggest that the user of the algorithm is irrelevant for purposes of accountability. If the goal is to reduce discrepancies and the

[112] Deirdre K Mulligan and Kenneth A Bamberger "Saving Governance by Design" (2018) 106 Calif L Rev 697 at 711
[113] Cohen and Graver, above n 107, at "455.
[114] Sheppard, above n 81, at 42
[115] Citron, above n 111, at 1272.
[116] Brauneis and Goodman, above n 11, at 127.
[117] Herschel Prins *Will They Do it Again? Risk Assessment and Management in Criminal Justice and Psychiatry* (London, Routledge, 1999) at [2]
[118] Oswald, Grace, Urwin and Barnes, above n 17, at 237; Corbett-Davies, Goel and González-Bailón, above n 17.

discrepancies arise from different people's interpretation of risk, the different people using the tool should no longer matter.

The same argument can be made in relation to reducing error. Errors are sometimes made in risk assessments when humans make judgments.[119] If algorithms can remove human frailty and produce more accurate assessments about risk,[120] it must be the algorithm itself that is significant rather than the individual using it. Cumulatively factors of consistency and error reduction suggest that the actor should go beyond just the user.

If the user of the algorithm is not necessarily the best actor to render account, the issue becomes identifying which party is best to pursue. An analogous case is automated weapons systems. Placing a human into the process to oversee the technology does not completely answer all questions related to responsibility.[121] Having a human involved as an operator of these systems does not cover all possible avenues for responsibility, as developers may also be implicated.[122] A framework of responsibility and accountability thus needs to recognise the potential breadth of actors who have accountability for the use of algorithms to ensure fairness.[123] In this case, we would have to look for another actor to interrogate.

Instead of focusing on one decision maker or entity to hold accountable, there may instead be multiple responsible parties.[124] Parties could include the creator of the algorithm, the user or the person who inputs the information, or the department who have decided to adopt the technology. There may need to be some accountability for those not just operating the algorithms, because unless that person approves every new decision and change in how the algorithm makes predictions as it learns from data, that person does not necessarily have control of the process itself.[125] Requiring the operator to approve every change in calculation on the part of the algorithm would be inefficient and would defeat the purpose of having a machine learning algorithm.[126] There is however pre-programming control that may give rise to an accountability responsibility.[127] Autonomous cars may be a good analogy. There are debates over how liability should be spread: software developers (of which there may be many), the manufacturer, or the owner of the car.[128] All of these people may have a contribution

[119] Gifford, above n 100, at 32.
[120] At 32.
[121] Calo, above n 47, at 416
[122] At 416
[123] At 416
[124] Frank Pasquale "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society" (2017) 78 *Ohio St LJ* 1243 at 1254.
[125] Karni Chagal-Feferkorn "The Reasonable Algorithm" (2018) Ill. JL Tech and Pol'y 111 at 133
[126] At 133
[127] At 133
[128] Emma L Flett and Jennifer F Wilson "Artificial Intelligence: Is Johnny 5 Alive? Key Bits and Bytes from the UK's Robotics and Artificial Intelligence Inquiry" (2017) 23(3) *CTLR* 72 at 73.

and all should be able to be held accountable so as not to avoid responsibility when things go wrong.

One such other party might be the developers of technology. Those responsible for the input have some responsibility for the output.[129] Where an algorithm is complex, such as machine learning algorithms, the designer may actually have more need for accountability because the way they have created it means that others cannot meaningfully understand it.[130] The creator is knowledgeable and has expertise so are most capable of being held accountable for the harms that the algorithm creates.[131] While machine learning can create new ways of interpreting data, the overall task or goal that it is trying to achieve must be programmed.[132] Accordingly, there is still a role for responsibility of developers and programmers.

The government may lack expertise to address challenges posed by algorithms, meaning they are reliant on creators of algorithms.[133] Because of this reliance, accountability for companies creating these algorithms is important. There is a loss of accountability where the person implementing the algorithm and making decisions on the basis of its recommendation does not understand how it works and the public cannot ask questions of the operator and so fairness is lost.[134] Because of the lack of information that the operator and government have, creators of algorithms may be a beneficial actor to pursue when it comes to accountability because they have expert knowledge and can explain how the algorithm works.

Requiring programmers to be accountable for the algorithms they produce is not without problems. First, after the author finishes writing the code it becomes independent of that person when put into use.[135] The way it is used in context to make decisions has some impact on whether the algorithm is used correctly and whether the actions of the user can also be held to account. Additionally, those who write codes often do not write the whole programme, just sections of it meaning that they do not see the whole picture.[136] Unless a specific part of the code was identified as the issue (which may be difficult), there may be problems in identifying an individual to ask questions of and demand answers from. Nevertheless, the company or wider entity creating the code may be able to step in, as government departments often do to be the actor in this scenario.

[129] At 73.
[130] Martin, above n 20, at 10.
[131] At 10.
[132] Sheppard, above n 81, at 44
[133] Calo, above n 47, at 428
[134] Brauneis and Goodman, above n 11, at 109
[135] Martin, above n 20, at 6.
[136] Chagal-Feferkorn, above n 125, at 133

Overall, the use of algorithmic technology disrupts understandings of accountability when working out the appropriate actor to hold accountable. The user or operator of the algorithm may defer to the algorithm making it hard to say that they have made a decision, meaning the developers of the algorithm may be more appropriate to hold accountable. Furthermore, even if users do not automatically defer to the algorithm's assessment, the operator or responsible minister may not have sufficient knowledge about how the algorithm works, meaning they may provide inadequate answers to the forum. As a result, accountability is not effective because answers are not accurate or helpful. In turn, sanctions may be based on incorrect or misunderstood explanations, reducing the efficacy of the consequence and opening up the possibility that similar mistakes will be repeated.

Not only are there problems with identifying the actor, there are also questions as to whether existing forums are sufficient. Forums may not have the appropriate technical expertise to hold algorithms accountable and regulate the technology.[137] They therefore may not be able to effectively ask questions and demand answers, which is central to be able to hold the actor (whoever they may be) accountable.

Common forums are courts and political superiors, however the public also have a more indirect role by being able to put political pressure on actors, and by bringing their cases to the courts if necessary. [138] The public additionally have a role in holding those who have been given delegated power to account as those who elected representatives.[139] Software can alter these relationships.[140] These forums are often not used to dealing with advanced technology such as algorithms so our usual avenues for scrutinizing decision making of actors is limited.[141] There are a number of general challenges that face all forums when it comes to human controlled machine learning.

1 The code may not be transparent

First, transparency creates issues for forums being able to effectively hold actors to account. Transparency assists with accountability as it exposes what might otherwise remain hidden and allows those not involved in the process to be confident of what delegates are doing.[142] If forums cannot see or discuss risk assessment algorithms, accountability is lost because they do not have all the information to be able to ask questions.[143] For the public, this

[137] Mulligan and Bamberger, above n 112, at 698
[138] Kroll, Huey, Barocas, Felten, Reidenberg, Robinson and Yu, above n 48, at 702
[139] Sinclair, above n 76, at 225.
[140] Kroll, Huey, Barocas, Felten, Reidenberg, Robinson and Yu, above n 48, at 702
[141] At 703
[142] John Roberts "No One is Perfect: the Limits of Transparency and an ethic for 'intelligent' accountability" (2009) 34 *Accounting, Organisations and Society* 957 at 957.
[143] Citron, above n 111, at 1254

relates not only to the content of the algorithm but even if it exists if it is not common knowledge outside of government that algorithms are being used. It was not known that Immigration was using the spreadsheet tool and so no concern from the public could be raised until it was exposed. Additionally, the fact the government is now looking into the use of algorithms is concerning as it might suggest that even officials are unaware of the extent of algorithm use and where they are employed in decision making. If our representatives do not even know whether algorithms are part of a decision, how is the public expected to know? Transparency alone probably will not be enough to ensure accountability, however it does allow for more effective auditing and allows meaningful legal challenges to their use, as well as increasing understanding.[144] Thus the first challenge for forums is determining whether algorithms are in use in the first place.

The second issue with transparency is understanding. Concerns about transparency, or lack thereof, have implications for the public's ability to understand and thus challenge the use of algorithms.[145] Lack of transparency also makes it difficult for experts to assess the tool.[146] Codes are not always made public for scrutiny. Developers of risk assessment algorithms in the United States have been unwilling to share how the algorithm works, and consequently there are concerns about transparency and accountability.[147] The company who developed the COMPAS software has not disclosed how the programme works and how decisions about risk levels are made.[148] For this reason, it is hard for forums to determine what factors have been used and weighted in the determination.[149] If the learned formula is not disclosed, there are the same problems as an individual's decision where the reasons are effectively inscrutable. If forums cannot see what the tool uses and how it makes its assessment then they cannot challenge its use.

Furthermore, disclosing code of machine learning algorithms is particularly difficult because the rule is created by analysis of the specific data rather than by programming.[150] The equation that turns data points into a risk level can be hard to understand or even determine,[151] especially since the algorithm itself decides in machine learning how factors are weighted and what interaction between the factors leads to the most accurate results.[152] While we may be

---

[144] Danielle Kehl, Priscilla Guo and Samuel Kessler "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing" (July 2017) Harvard University <www.cyber.harvard.edu> at 33.
[145] Oswald, Grace, Urwin and Barnes, above n 17, at 237
[146] Kehl, Guo and Kessler, above n 144, at 28.
[147] Gifford, above n 100, at 33.
[148] At 32.
[149] Angwin, Larson, Mattu and Kirchner, above n 15.
[150] Kroll, Huey, Barocas, Felten, Reidenberg, Robinson and Yu, above n 48, at 638.
[151] Angwin, Larson, Mattu and Kirchner, above n 15
[152] Stevenson, above n 50, at 10.

able to see the data inputs, how the programme combines them and weights them is more complex and is what leads to lack of transparency.[153]

It is important to keep in mind though that while transparency is important to ensure accountability, it is not the sole feature that is required. Transparency is useful in aiding in constraining power however it is also only one part of accountability and should not be used as a stand in for the broader concept.[154] Releasing large amounts of information may not help the public understand or hold actors accountable. Simply providing information does not necessarily mean scrutiny by a forum.[155] You can reveal something without necessarily talking about it and demanding justifications or imposing sanctions. Transparency alone is not sufficient because those receiving the information do not necessarily have the in depth knowledge about what the information means or contextual markers.[156] Judges often hear cases where they do not have all the information, whether as a result of gaps in the evidential record or because evidence has been excluded.[157] Consequently the idea that oversight requires full information can be rebutted.[158] Therefore, even without full knowledge of the contents of algorithms forums, courts especially can still make a principled and reasoned judgment about whether there needs to be sanctions imposed for the sake of accountability.

Without transparency however, the observer would be reliant on what they are told by others without the supporting information.[159] It also provides an opportunity for subsequent steps to occur through debate and imposing sanctions on the basis of what is shown by the information disclosed. For this reason the debate stage is important, so that information behind what has been disclosed can be probed. Still, initial transparency is important. Transparency is the beginning of the process. Take the immigration case, nothing could begin to be done until the public found out about the potential issue.

Overall, machine learning algorithms may be ill-suited to many forums of accountability. The challenges arise from lack of transparency as a result of source codes not being disclosed and the complexity of how variables interact if the code is released.

---

[153] Oswald, Grace, Urwin and Barnes, above n 17, at 233.
[154] Roberts, above n 142, at 968.
[155] Bovens, above n 29, at 453.
[156] Roberts, above n 142, at 964
[157] 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 L & Tech Rev* 18 at 43
[158] Kroll, Huey, Barocas, Felten, Reidenberg, Robinson and Yu, above n 48, at 705.
[159] Roberts, above n 142, at 964

2 *Data disclosed may be too complex*

Complexity poses another barrier for forums to effectively impose accountability for the use of machine learning algorithms. Regardless of whether there is complete transparency, understanding the technology and how it works is difficult. Even if full disclosure is possible, there is likely to be too much code and data to be able to effectively analyze.

For example, the HART model has over 4.2 million decision points. [160] While it may be possible to disclose all of these, in reality it would be too difficult for an individual, particularly a member of the general public, to understand and evaluate all of them.[161] Thus both lay and expert forums would unlikely be able to look at all of the variables. Additionally, although all the inputs are known, the interaction of the information may be too difficult to interpret.[162] The volume of data and complexity of how these data points interact therefore create problems even for expert forums, let alone forums with less technical expertise like the public and political superiors. The debate stage which is important for accountability is therefore severely hampered by sheer volume and complexity of information even if there is full transparency and high levels of expertise.

3 *Forums may not have requisite expertise*

Even if there was full disclosure and not a large amount of data, most forums would likely lack expertise to interpret that data. Consequently, issues arise in the forum being able to ask meaningful questions as they might not understand how the algorithm works.

Courts are a specific forum that may not have the expertise to analyse the data and technological components, as a judge is unlikely to have detailed knowledge of computer programming or coding. For this reason, they may not fully appreciate how the algorithm works and how it contributes to the decision that is being made. In *Wisconsin v Loomis* it was noted that the Court struggled to understand the algorithm.[163] Given that judges are not unintelligent individuals, this highlights the difficulties that lay people, and even politicians may have in understanding algorithms and thus holding individuals to account for their use.

Ministers too may lack expertise to be able to ask appropriate questions of the department. Again, the Immigration example is a good starting point. In the example, the tool was a simple spreadsheet. Most individuals are familiar with spreadsheets, meaning that the

[160] Oswald, Grace, Urwin and Barnes, above n 17, at 234
[161] At 234
[162] National Science and Technology Council, above n 51, at 9.
[163] *Wisconsin v Loomis,* above n 102, at [133].

Minister was able to understand and appropriately respond. Had the tool been more complex, Mr Lees-Galloway may not known what to ask.

The public too face challenges related to expertise when acting as a forum. Since the public might not have the specialist knowledge to interrogate the information disclosed, they will still be reliant on what they are being told, placing trust in those with delegated power.[164] People would have to have the skills to understand the data if they are to effectively question it and to maximise accountability, however high level programming is not a skill that many members of the public have. As already discussed, source codes of machine learning algorithms are extremely complex and hard to understand even for experts. In this way, the public as a forum may be ill equipped to hold the actors accountable, meaning there are no benefits for accountability in using the public as a forum for complex algorithms.[165] The public may be able to initiate a general discussion but where there may be a challenge is in the follow up and determining whether answers are accurate and satisfactory. Thus accountability is frustrated given that lack of expertise means that forums may be ill-equipped to determine whether the actor's answers are satisfactory and may not even know what relevant questions to ask in the first place.

However, forums may be able to rely on those who do have expertise. Experts may be used in court for judges to be able to understand how algorithms work, and parties like the creators and programmers of the code can be called as witnesses to further increase understanding. Judicial review for example involves a judge looking at decisions made by public bodies to determine if they have acted within their powers.[166] However, algorithms mean that experts must take on some of this burden as courts may no longer be able to determine how a decision has been made and whether legal rules have been followed by the algorithm if it formed a significant part of the decision.[167] While this may assist in laying a base knowledge, there may be issues later in the process when the judge is deliberating and making their decision if questions arise that were not adequately addressed by experts during the trial.

Ministers may also have technical advisers to assist them, and as complex machine learning algorithms become more common this may also become more common. Thus like courts, they may increasingly rely on experts.

Likewise, just because the public are not experts on coding or algorithms does not mean that public accountability is impossible for complex algorithms.[168] The public can rely on

[164] Roberts, above n 142, at 966
[165] Edwards and Veale, above n 157, at 41.
[166] Kroll, Huey, Barocas, Felten, Reidenberg, Robinson and Yu, above n 48, at 703
[167] At 703.
[168] Edwards and Veale, above n 157, at 43

experts to inform them and to enable them to ask questions and assess the accuracy and sufficiency of answers they receive. The public is used to hearing from experts in the media about complex issues that they do not have personal expertise in, so it is likely this would be sufficient to enable the public to ask relevant questions and determine whether the answers are acceptable. Nevertheless, most other policy and issues that the public are used to interpreting and assessing are arguably not as complex as machine learning algorithms. Background knowledge might be important to understanding implications and oversimplification in pursuit of public understanding may be dangerous if it obscures real issues.

Still, the early stages of the Immigration example may serve to show the public is an effective forum for algorithmic accountability, to at least initiate discussion. In the early stages it was thought that the tool that Immigration was using was highly complex, yet despite not necessarily understanding it, the public was still able to hold Immigration New Zealand accountable, as they had to answer questions both in the public forum through the media and respond to Mr Lees-Galloway. This highlights the idea that the public does not necessarily need to understand or even see the whole tool or algorithm to demand answers. Thus full understanding and transparency is not required. The initial discussion in the public forum also led to the initiation of a closer assessment of the use of algorithms in the public sector, so the more detailed and technical issues may be able to be more appropriately addressed in that context.

Overall, there may be a lack of expertise on the part of forums, impacting their ability to ask probing questions and assess responses from actors. To some degree this may be helped by the fact that forums can obtain assistance from experts.

4 Absence of reasons given by the algorithm

Another issue is lack of reasons. Algorithms are potentially more transparent than human decision makers because we can see the factors that are inputted. However, where humans may have the edge in decision making and allowing forums to effectively demand accountability is that algorithms do not provide reasons.

It is rare that algorithms give reasons for their decisions, for example the programme AlphaGo beat the best human Go player but did so by making a strange move that could not be explained by programmers.[169] Lack of explanation, while inconsequential in a game, are important when there are more important issues such as liberty at stake,[170] as is the case with risk assessment tools. Some programmes such as LIME try to explain their decisions by

[169] Flett and Wilson, above n 128, at 72.
[170] At 72.

articulating the main factors that were likely to have influencing the classification.[171] While this may help, the actual process for the decision is still not explained since just because a factor has been given the most weight does not explain how it is related to other data that may have also had an impact on the decision and it may still only be a small part in the overall decision.[172] Additionally, in complex decisions where the decision could easily go one way or another there are a number of factors that may not have a significant weight that nevertheless are what has tipped the scale.[173] Thus forums face significant barriers when trying to scrutinise reasons for decisions while trying to hold actors accountable.

Forums such as courts may find it difficult to hold actors to account if reasons are not given, as any errors will be harder to review.[174] Decisions made by judges and other public decision makers are less opaque than algorithmic decisions as reasons are stated and their decisions may be questioned later.[175] So while the algorithm is accurate in finding correlations, it cannot explain why there is a causal connection.[176] The lack of reasoning presents challenges for the usual forums because all we have to go on is the inputs.

Courts rely on reasons of decision makers to review decisions. These allow parties to see that there has been careful consideration and so are evidence that the issue has been well thought out. [177] Absence of these may make it harder to determine whether there was an error in the application of the decision making power.[178] Despite the fact that transparency is not necessarily helpful in all situations related to algorithms, courts and other bodies that ensure responsibility such as the Ombudsman require some form of explanation or transparency to be able to do their jobs.[179] Other forums like ministers and the public also require some kind of explanation for the exercise of decision making power to pass judgment on whether it was justified.

Even with oversight from a human, that human does not know why the algorithm generated a particular risk score so if it is relied on (which is likely due to automation bias) no meaningful reasons can be given. Without these meaningful reasons, forums' questions may be superficial or be based on misinterpretation or misunderstanding of information.

---

[171] Sheppard, above n 81, at 49
[172] At 50
[173] At 50.
[174] Kendrick Lo "When Efficiency Calls: Rethinking the Obligation to Provide Reasons for Administrative Decisions" (2018) 43 *Queen's LJ* 325 at 326
[175] Pasquale and Cashwell, above n 80, at 66.
[176] Calo, above n 47, at 414
[177] Lo, above n 174, at 352.
[178] Lo, above n 174, at 352.
[179] Edwards and Veale, above n 157, at 40

5 Deference

While actors face machine bias, it is also present in forums. Judges when reviewing decisions that rely on algorithms may be unwilling to challenge algorithmic decisions due to the perception that algorithms are more objective and therefore superior to human judgment.[180] Likewise, ministers and the public may think that there is no issue to be addressed or be unwilling to impose sanctions when the algorithm has had a significant impact on the process. The positives of risk assessment algorithms in reducing bias and being more accurate may mean that judges, politicians, and the public do not feel as though they can oppose what is perceived to be a highly accurate and robust tool.

6 Judicial review as an example of challenges for forums

To illustrate how forums and their existing accountability mechanisms may be ill equipped to deal with increasingly complex algorithms, we can use the example of judicial review. Judicial review may be limited or rendered ineffective because judges might not be able to see how the programme made the particular decision.[181]

When asking whether the decision maker considered irrelevant factors algorithms may present problems. Aside from issues of knowing exactly what factored into an algorithm's assessment, if the algorithm was unreliable it would be hard to see how it would be justifiable as a relevant consideration.[182] Without knowing the formula (as would be the case in machine learning algorithms), determining whether it was reliable or not may be a challenge. But it may be that results of an algorithm are determined to be relevant if they are deemed to be more reliable than other forms of risk assessment.[183]

Furthermore, if nuanced factors that a human decision maker could easily take into account such as family or employment are relevant but not part of the algorithm (or if we can't work out whether they were part of the decision), this may too lead to a successful challenge.[184] Inputs can be clearly seen and what is entered into the algorithm but these are all that is taken into account by the programme. It may be that this means it is easier to see what is factored into the decision. But this also means that other relevant factors may not be considered and so the decision is not valid. Mitigating this concern is the fact that ultimate human control would at least know that certain factors were not an input, even if they didn't know how the algorithm calculated the risk score. They could therefore manually factor those pieces of information in.

---

[180] Brauneis and Goodman, above n 11, at 126.
[181] Citron, above n 111, at 1298
[182] Oswald, Grace, Urwin and Barnes, above n 17, at 239.
[183] At 239
[184] At 240

Additionally, some inputs may also be proxies for other factors which may be irrelevant considerations.[185] Despite algorithms assisting in reducing bias, they are not immune because inputs such as education may be correlated to race.[186] Consequently, the algorithm may still reflect racial inequalities which are embedded in society and so will make assessments that perpetuate racial discrepancies in criminal justice outcomes. In fact, the Waitangi Tribunal report dealt with issues about including race as an input into algorithmic decision-making,[187] albeit in relation to formulas rather than machine learning. These indirect relationships might be overlooked or be unclear to forums who accept the inputs at their face value. If it is unclear how these are weighted in the algorithm's assessment it may be difficult for a claimant to say whether and how these influenced the final decision.[188] Consequently, the court is prevented from being sure what was considered and thus whether the actor exercised their power appropriately.

Natural justice may also be affected. Natural justice requires decision makers give reasons and are transparent about their processes that would allow the decision to be challenged if necessary.[189] As previously discussed, there are challenges with transparency as they relate to algorithms so algorithms as a decision making tool may be unacceptable from a natural justice perspective. Further, the decision may not be explicable because the programme does not give reasons for its decision and cannot explain why a decision has been made.[190] Coupled with aforementioned problems of understanding and the fact that even if all inputs are disclosed there may be too much information for humans to successfully analyse, this may be a breach of natural justice rights.[191]

It may also be that we do not know a lot about how effective they actually are yet and so proportionality tests involved in judicial review would be hampered.[192] This concern also applies to other legal tests, particularly in Bill of Rights cases where rights are often engaged in the criminal justice context.[193] In this context, the New Zealand Bill of Rights Act requires that any limits on rights be "demonstrably justified in a free and democratic society".[194] This test requires overall proportionality.[195] This would mean that not knowing the efficacy of the

[185] Kehl, Guo and Kessler, above n 144, at 28
[186] Hamilton, above n 10, at 262.
[187] Waitangi Tribunal, above n 5.
[188] Kehl, Guo and Kessler, above n 144, at 28
[189] Oswald, Grace, Urwin and Barnes, above n 17, at 241.
[190] At 241
[191] At 242
[192] At 242
[193] BORA readings
[194] New Zealand Bill of Rights Act 1990, s 5.
[195] *R v Oakes* [1986] 1 SCR 103 at [70]; *R v Hansen* [2007] NZSC 7, at [64].

algorithm or what their process is makes it is hard to determine whether the algorithm is the least intrusive means necessary to achieve the goals of the decision maker.

Due to lack of expertise to ask the right questions and complexity of the machine learning algorithms, other forums may become more common in this area, like professional bodies and experts. Often we must delegate tasks to those who have specialist expertise and so we need a way to ensure that these people are responsible even if we do not fully understand their tasks as they are still ultimately responsible to political representatives.[196] In some circumstances, this may require review by other experts in the field, rather than by political figures as this is the only way that their actions can be fully interrogated and assessed.[197] To ensure efficacy of external review of decisions made in an area which is highly specialised, expert panels or bodies may have to be set up to ensure that they are ultimately responsible to the public.[198] An independent and expert body is likely needed to assess and monitor the use of algorithms because of the complexity of programming and statistical information used.[199] Again however there may be problems of accessibility for others to have oversight to be sure that this forum is adequately performing their function.

D Autonomous machine learning

A step further would be to use machine learning algorithms unsupervised by humans and apply them with no human input or override. As mentioned, it would be unlikely that this would happen, at least not in the foreseeable future.

If we were to adopt autonomous machine learning however, who would we look to as an actor? Since there is no human user, there would be no issues of whether that person is appropriate as an actor to address a forum's concerns. Using algorithms alone however could raise interesting questions of whether they could be treated as fully independent of humans.[200] Instead, the actor and "user" could be the algorithm itself. In order to be able to satisfy requirements of providing a satisfactory account to a forum, the algorithm would have to have some kind of responsive capability. To be able to do this, the algorithm would need to be able to give reasons for its decision and justify it otherwise there is no accounting. Moreover, there could be problems with these reasons if they do not explain how the factors interact or if they only cover the main reasons for the decision.[201] As discussed above there are a number of challenges with the existing programming that provides reasons in machine learning. As it

[196] Mulgan, above n 31, at 558.
[197] At 559.
[198] At 559.
[199] Oswald, Grace, Urwin and Barnes, above n 17, at 241
[200] Giuffrida, Lederer and Vermeys, above n 103, at 763-764.
[201] Sheppard, above n 81, at 50

stands therefore, it is likely that holding the algorithm itself to account would be impractical if we want detailed and meaningful responses.

For those detailed and meaningful responses, it is likely that a human would need to be involved to act as a "translator" for the algorithm. There may still be a role for developers if the algorithm was inaccurate as it would be a consequence of faulty programming. However, the use of programmers as an actor runs into many of the same problems as above. They may not even be able to fully explain the machine's decision due to complexity of the formula and problems with transparency due to volume of information. Political superiors may also be an actor if explaining why the technology was adopted but not how it works necessarily.

Forums may be significantly frustrated in their function if we adopted autonomous machine learning for risk assessment decisions. Forums to pass judgment and impose consequences require substantive information. Because of the limited nature of the actor's ability to give responses, it is unlikely that the forum would be able to take any meaningful action based on robust explanations.

As discussed, we can see what inputs are used,[202] so if there is no human element of whether other factors were considered, judicial review might be able to look at these and come to an assessment. The court may be able to make this assessment because it will be easier to see whether relevant or irrelevant factors were used in the decision.[203] However, forums also need to be able to see how the factors that are used are turned into a decision as the raw data itself is not the problem, it is how a risk score is determined from this information. Thus issues of lack of transparency about the actual equation that the algorithm has generated remain an issue. Since this remains an issue, so too do the related issues of complexity of code, volume of code and lack of expertise.

VI Conclusion

I have identified several themes that emerge from looking at the issues with using algorithms in risk assessment decisions. These issues feed into the difficulty in ensuring that decisions made using technology are subject to accountability mechanisms. Accountability for the use of algorithms is hampered by the fact that this technology disrupts conventional structures of accountability. They do so by making it potentially unclear who the actor who we should be holding accountable is as there may be multiple actors who have a hand in the process that leads to a questionable outcome. There is also the question of responsibility distribution among these actors, particularly if users of the tools do not understand how the tools work.

[202] Oswald, Grace, Urwin and Barnes, above n 17, at 233
[203] At 240

Additionally, forums that are commonly used may be ill equipped to ask questions and demand answers which is critical for effective accountability.

Challenges that accountability structures face in relation to new machine learning algorithms used in risk assessment are that it may be hard to identify the appropriate actor to hold accountable. Across actors, there are potentially many parties we can demand account from. Among these, options may not have meaningful control over the decision or lack expertise. As a result, it may be ineffective to scrutinize them and ask them to explain the part of the decision making process the algorithm assists with. In these situations, it may be more effective to hold the creators of these algorithms accountable as they do have specialist knowledge.

Usual forums relied on to ensure accountability for criminal justice decisions may also need revising. Across forums there is difficulty in having the appropriate expertise and having to rely on external experts. There is also an issue of transparency with machine learning (and even potentially some formulas if they are not disclosed) where we do not know how the factors interact. Even if we did, the sheer volume of data is also a barrier to effective scrutiny, since all information may not be able to be realistically understood and interpreted by forums. Understanding algorithms requires specialist knowledge and the technology used is often complex, so even experts may find it difficult to effectively question risk assessment algorithms.

Thus the biggest challenge is issues with understanding for both actors and forums. Alternatives should be found to address this so accountability can successfully intervene where it needs to and so mistakes are prevented in the future. These issues should be thought about and solutions developed. Such solutions are beyond the scope of this paper, however they are critical to ensure criminal justice decisions continue to be subject to accountability even when they are made in reliance on technology.

VII Bibliography

A Legislation

New Zealand Bill of Rights Act 1990.

Parole Act 2002.

Sentencing Act 2002.

B Cases

R v AM [2010] NZCA 114.

R v Hansen [2007] NZSC 7.

R v Oakes [1986] 1 SCR 103.

Wisconsin v Loomis 68 2015AP157-CR (Wis 2016).

C Reports

Leon Bakker, David Riley, James O'Malley "Risk of Reconviction: Statistical Models Predicting Four Types of Re-Offending" (Department of Corrections, 1999).

Department of Corrections Policy, Strategy and Research Group "Over-representation of Māori in the criminal justice system: An exploratory report" (Department of Corrections, September 2007).

National Science and Technology Council "Preparing for the Future of Artificial Intelligence" (United States Government, October 2012).

Lyn Provost "Department of Corrections: Managing off enders to reduce reoffending" (Office of the Auditor General, December 2013).

Waitangi Tribunal *The Offender Assessment Policies Report* (Wai 1024, 2005).

D Books

Hazel Kemshall *Understanding Risk in Criminal Justice* (Maidenhead, Open University Press, 2005).

Geoffrey Palmer and Matthew Palmer *Bridled Power* (4th ed, Melbourne, Oxford University Press, 2004).

Herschel Prins *Will They Do it Again?: Risk Assessment and Management in Criminal Justice and Psychiatry* (London, Routledge, 2002).

Raul Rishworth "The New Zealand Bill of Rights" in Grant Huscroft, Paul Rishworth, Richard Mahoney, and Scott Optican (eds) *The New Zealand Bill of Rights* (Oxford University Press, Oxford, 2014) 1.

E Articles

Richard Berk and Jordan Hyatt "Machine Learning Forecasts of Risk to Inform Sentencing Decisions" (2015) 27(4) *Federal Sentencing Reporter* 222.

Mark Bovens "Analysing and Assessing Accountability: A Conceptual Framework" (2007) 13(4) European Law Journal 447.

Robert Brauneis and Ellen P Goodman "Algorithmic Transparency for the Smart City" (2018) 20 *Yale LJ & Tech* 103.

Cary Coglianese and David Lehr "Regulating By Robot: Administrative Decision Making in the Machine Learning Era" (2017) 105 *Geo LJ* 1147.

Glenn Cohen and Harry S Graver "Cops, Docs, and Code: A Dialogue Between Big Data in Health Care and Predictive Policing" (2017) 51 *UC Davis L Rev* 437.

Ryan Calo "Artificial Intelligence Policy: A Primer and Roadmap" (2014) 51 *UC Davis L Rev* 399.

Alyssa M Carlson "The Need for Transparency in the Age of Predictive Sentencing Algorithms" (2017) 103 *Iowa Law Review* 303.

Karni Chagal-Feferkorn "The Reasonable Algorithm" (2018) Ill. JL Tech and Pol'y 111.

Anna Chalton "Rape Myths and Invisible Crime: The Use of Actuarial Tools to Predict Sexual Recidivism" (2014) 5 PILJNZ 112.

Danielle Keats Citron "Technological Due Process" (2008) 85(6) *Washington University Law Review* 1249.

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 L & Tech Rev* 18.

Emma L Flett and Jennifer F Wilson "Artificial Intelligence: Is Johnny 5 Alive? Key Bits and Bytes from the UK's Robotics and Artificial Intelligence Inquiry" (2017) 23(3) *CTLR* 72

Richard Gifford “Legal Technology: Criminal Justice Algorithms: AI in the Courtroom” (2018) 38(1) *The Proctor* 32.

Iria Giuffrida, Fredric Lederer and Nicholas Vermeys “A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and other Technologies will Affect the Law” (2018) 68 *Case W Res* 747.

Melissa Hamilton “Risk-Needs Assessment: Constitutional and Ethical Challenges” (2014) 52 *American Criminal Law Review* 231.

Eric Holder, U.S. Attorney General “Speech Presented at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference, Philadelphia, PA” (2015) 27(4) *Federal Sentencing Reporter* 252.

Danielle Kehl, Priscilla Guo and Samuel Kessler “Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing” (July 2017) Harvard University <www.cyber.harvard.edu>.

Joshua A Kroll, Joanna Huey, Solon Barocas, Edward W Felten, Joel R Reidenberg, David G Robinson and Harlan Yu “Accountable Algorithms” (2017) 165 *U Pa L Rev* 633.

Kendrick Lo “When Efficiency Calls: Rethinking the Obligation to Provide Reasons for Administrative Decisions” (2018) 43 *Queen’s LJ* 325.

Kirsten Martin “Ethical Implications and Accountability of Algorithms” (2018) *Journal of Business Ethics* at 7.

John Monahan and Jennifer L Skeem “Risk Assessment in Criminal Sentencing” (2015) *Annual Review of Clinical Psychology* (forthcoming).

Richard Mulgan “Accountability: An ever expanding concept?” (2000) 78(3) *Public Administration* 555.

Deirdre K Mulligan and Kenneth A Bamberger “Saving Governance by Design” (2018) 106 Calif L Rev 697

Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey C. Barnes “Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental’ proportionality” (2018) 27(2) *Information & Communications Technology Law* 223.

Frank Pasquale and Glyn Cashwell "Prediction, Persuasion and the Jurisprudence of Behaviourism" (2018) 68 U Toronto JL 63.

Frank Pasquale "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society" (2017) 78 *Ohio St LJ* 1243.

John Roberts "No One is Perfect: the Limits of Transparency and an ethic for 'intelligent' accountability" (2009) 34 *Accounting, Organisations and Society* 957.

Lauren Henry Scholz "Algorithmic Contracts" (2017) 20 *Stand Tech L Rev* 128.

Brian Sheppard "Warming up to Inscrutability: How Technology Could Challenge Our Concept of Law" (2018) 68 *U Toronto LJ* 36

Amanda Sinclair "The Chameleon of Accountability: Forms and Discourses" (1995) 20(2/3) *Accounting, Organisations and Society* 219.

Megan Stevenson "Assessing Risk Assessment in Action" (2018) 103 *Minnesota Law Review*.

Kaare Strøm "Delegation and accountability in parliamentary democracies" (2000) 37 *European Journal of Political Research* 261.

Carla Swansburg "Artificial intelligence and Machine Learning in Law: The Implications of Lawyers' Professional Responsibilities for Practice Innovation" (2018) 60 *CBLJ* 385

Dennis F Thompson "Moral Responsibility of Public Officials: The Problem of Many Hands" (1980) 74(4) *The American Political Science Review* 905.

F Internet materials

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner "Machine Bias" (23 May 2016) ProPublica <www.propublica.org>.

BBC "Police warned about using algorithms to decide who's locked up" (16 November 2017) BBC <www.bbc.com>.

Gill Bonnett "Immigration New Zealand using data system to predict likely trouble makers" (5 April 2018) Radio New Zealand <www.radionz.co.nz>

Sam Corbett-Davies, Sharad Goel and Sandra González-Bailón "Even Imperfect Algorithms Can Improve the Criminal Justice System" (20 December 2017) The New York Times <www.nytimes.com>.

Nicholas Diakopoulos "How to Hold Governments Accountable for the Algorithms They Use" (2016) Slate <www.slate.com>.

Tom Pullar-Strecker "'Urgent' algorithm stocktake to show how government uses our data" (24 May 2018) Stuff <www.stuff.co.nz>.

Radio New Zealand "Immigration Minister puts controversial profiling programme on hold" (9 April 2018) Radio New Zealand <www.radionz.co.nz>

Lincoln Tan "Immigration NZ's data profiling 'illegal' critics say" (5 April 2018) New Zealand Herald <www.nzherald.co.nz>.

Jason Tashea "Risk-assessment algorithms challenged in bail, sentencing and parole decisions" (March 2017) ABA Journal <www.abajournal.com>.