Seeing Eye to AI:
A Modest Case for an Algorithmic Sentencing System in New Zealand

Nur Syairah Mohd Nizam

A dissertation submitted in partial fulfilment of the degree of Bachelor or Laws (Honours) at the University of Otago – Te Whare Wānanga o Otāgo, Dunedin, New Zealand.

7 October 2021

ACKNOWLEDGEMENTS

To my supervisor, Professor Colin Gavaghan,
for providing invaluable guidance and feedback.
I would also like to thank you for sparking my interest in the area of emerging technologies,
your passion and enthusiasm in class is infectious.

To the High Court of Sabah and Sarawak,
for taking the time to answer my questions on the AI sentencing system in Malaysia,
which was the inspiration behind this dissertation.

To Nur Awab, Hannah Shribman-Brawn and Nurain Nasuha,
for acting as my extra pair of eyes in proofreading this dissertation.

To Abdul Rahman,
for being my personal cheerleader throughout this whole dissertation process.

To my friends, inside and outside of law school,
for keeping me sane and grounded.

And finally, to my family, 8,891 kilometers away,
for the unwavering love and support.

TABLE OF CONTENTS

INTRODUCTION

Despite its increasing influence on our everyday life, many aspects of Artificial Intelligence (AI) remain a mystery to the human mind. Therefore, it is understandable that in the past, proposals to introduce AI into an already indeterminate and uncertain domain such as sentencing law are often met with scepticism and apprehension.[1] Such a proposal, at first instance, would seem to some, ambitious or even naive. However, considering that sentencing is a complex legal domain involving the balancing of various difficult considerations which has traditionally remained within the realm of human judges, such an assertion is reasonable. In this dissertation, however, I will make a modest case for the introduction of such technology into this domain and how best to regulate it. This dissertation contends that a new regulatory model in the form of an Independent Monitor would be adequate to address the harms associated with an algorithmic sentencing system.

My argument is modest in two respects. Firstly, this dissertation considers an incremental establishment of an algorithmic sentencing system that would operate alongside New Zealand's existing sentencing process. The predictions generated by the system in the form of a recommended sentence will serve nothing more than a non-binding recommendation. Secondly, the algorithmic sentencing system discussed in this dissertation does not aim to replicate the moral reasoning involved in the judicial decision-making process but rather to systemise past decisions made by human judges to inform the judge of how their colleagues in the past facing similar situations have reacted.

In chapter I, I will give a high-level summary of artificial intelligence and algorithms. In doing so, I will highlight the importance of viewing algorithms as sociotechnical systems rather than of "artifacts set apart".[2] This approach of viewing algorithm-driven tools in a meaningful context in which they will significantly operate will bring me to the examination of the

---

[1] Andrea Roth "Trial by Machine" (2016) 104(5) Geo LJ 1245; Ric Simmons "Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System" (2018) 52(2) UC Davis L Rev 1067; Michael E. Donohue "A Replacement for Justitia's Scales? Machine Learning's Role in Sentencing" (2019) 32(2) Harv JL & Tech 658.

[2] Carla L Reyes and Jeff Ward "Digging into Algorithms: Legal Ethics and Legal Access" (2020) 21 Nev LJ 325 at 343.

sentencing system in New Zealand. In this part of my dissertation, I will provide a brief overview of the sentencing process in New Zealand and its problems, by reference to four core judicial values that are relevant to the sentencing process – transparency, impartiality, consistency and efficiency. The purpose of doing so is to remind readers that when introducing a change to the status quo, the issue is comparative: How does this proposed change fare compared to the current system? This approach will encourage us to set realistic contextual goals, fueling robust discourse on regulatory issues relating to algorithmic decision-making. At the end of this chapter, I will also explain how an algorithmic sentencing system might improve the current sentencing process in New Zealand. Most importantly, I will argue that the need for an algorithmic sentencing system rests mainly on the goal of efficiency. While the system aims to achieve consistency, there are concerns that using an algorithmic sentencing system to improve consistency would not be desirable.

Chapter II will then proceed to give an account of what an algorithmic sentencing system in New Zealand could look like. It will draw from Vincent Chiao's recent work.[3] For guidance, I will also look to the AI-based sentencing system that the High Court of Sabah and Sarawak has recently introduced in its criminal justice system. After exploring what an algorithmic sentencing system in New Zealand could look like, I will then highlight the limitations of such a system in terms of transparency and impartiality. I want to point out here that such concerns are not unique to algorithmic sentencing systems alone, but also exist with other algorithmic systems used in other contexts. Finally, this chapter will explore the issue of judicial ambivalence as another concern relating to algorithmic sentencing systems.

Assuming the Ministry of Justice decided to introduce the algorithmic sentencing system in New Zealand courts, chapter III will explore the various recourses available in New Zealand's current legislative framework. These recourses are in the form of the new Privacy Act, the Official Information Act 1982 (OIA), the Local Government Official Information and Meetings Act 1987 (LGOIMA) and the Human Rights Act 1993 (HRA). I will also analyse the Algorithm Charter for Aotearoa New Zealand and evaluate its effectiveness in mitigating the

[3] Vincent Chiao "Predicting Proportionality: The Case for Algorithmic Sentencing" (2018) 37(3) Crim Just Ethics 238.

risks with algorithmic decisions. This chapter will conclude that existing legal avenues fail to appropriately respond to the concerns relating to algorithmic decisions mentioned in chapter II. However, I will submit that the Algorithm Charter would complement the new regulatory model proposed in the next chapter.

Chapter IV will look at the General Data Protection Regulation (GDPR) as a potential inspiration and show how the regime also has its flaws in addressing algorithmic transparency and bias. However, I contend that the European Commission's recent proposal for a "Regulation laying down harmonised rules on Artificial Intelligence" (AI Regulation) seems to be more promising in addressing harms associated with algorithmic decision-making. This chapter will also examine possible solutions to strengthen the current legislative framework. Finally, this chapter proposes a new regulatory model in the form of an Independent Monitor, which has been endorsed previously by several experts in this area.[4]

It is important to note that the use of an algorithmic sentencing system to recommend an overall sentence for an accused is not unheard of and even has real-life applications. On 19 February 2020, the Malaysian judiciary in Sabah and Sarawak introduced an AI-based system called AI in Court Sentencing (AICOS) to assist Magistrates in recommending an overall sentence for an accused.[5] The system is based on data analytic and machine learning, where a number of inputs consisting of aggravating and mitigating factors are inserted into the system.[6] The AI will then analyse past cases that have been programmed into its system and generate

---

[4] Colin Gavaghan, Alistair Knott, James Maclaurin, John Zerilli and Joy Liddicoat *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project* (Wellington, 2019) at 76 – 77; Andrew Tutt "An FDA for Algorithms" (2016) 69 Admin L Rev 83 at 117 – 122; Matthew U. Scherer "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) 29(2) Harv JL & Tech 353 at 393 – 398; Jacob Turner *Robot Rules: Regulating Artificial Intelligence* (Springer, Switzerland, 2019) at 254 – 262; and David Smith "The Citizen and the Automated State: Exploring the Implications of Algorithmic Decision-making in the New Zealand Public Sector" (Master's Thesis, Victoria University of Wellington, 2020).

[5] Olivia Miwill "Malaysian Judiciary Makes History, Uses AI in Sentencing" *New Straits Times* (online ed, Malaysia, 19 February 2020).

[6] Interview with AI Committee Members of Sabah and Sarawak Courts (the author, Zoom Meeting, 5 August 2021).

an output which, in this case, is the recommended sentence for the accused.[7] Currently, the system is only available for one offence; namely possession of drugs under section 12(2) of the Dangerous Drugs Act, 1952.[8] This pilot project is the first of its kind to develop an AI-based system that has similar functions to Sentencing Guidelines but is based on data analytics and machine learning. The primary objectives of the AI are to encourage consistency and parity in sentencing and to discourage hugely disparate sentences for similar offences.[9] While the AI provides recommendations based on selected criteria, the court must decide the most appropriate sentencing, according to the established principles of sentencing within the limits prescribed by the relevant legislation.[10]

[7] AI Committee Members of Sabah and Sarawak, above n 6.

[8] AI Committee Members of Sabah and Sarawak, above n 6.

[9] AI Committee Members of Sabah and Sarawak, above n 6.

[10] AI Committee Members of Sabah and Sarawak, above n 6.

CHAPTER I: SETTING THE SCENE

A What is Artificial Intelligence?

With its capacity to facilitate innovations from Google Maps to Netflix's recommendation algorithms, AI is rapidly influencing every aspect of our life. Despite its growing dominance, experts are unable to agree on a single definition of AI. This is mainly due to the abstract concept of intelligence itself.[11] Some experts broadly define AI as "a set of techniques aimed at approximating some aspect of human or animal cognition using machines".[12] Beyond that general definition, a widely agreed-upon definition of AI remains elusive. For the purpose of this dissertation, thinking about AI in terms of algorithms is a good starting point.

Algorithms are at the core of every AI system. Like a recipe used to prepare a meal, an algorithm is a "specific set of instructions used for calculating a function".[13] There are two classic types of algorithms: handcrafted algorithms, which are sometimes described as "the first wave of artificial intelligence", and machine learning algorithms, which form the second wave of AI, and are gaining momentum today.[14] Reyes and Ward would view these two types of algorithms as "computational components of algorithmic systems".[15] As I will discuss in more detail later, Reyes and Ward encourage a shift from viewing algorithms as independent or as "artifacts set apart" to viewing them as social technological systems, "set within and interacting with humans in social context".[16] I submit that, such a shift in perspective would enable deeper and more robust discourse on key design and regulatory questions involving algorithmic decision-making tools.

---

[11] Andreas Kaplan and Michael Haenlein "Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence" (2020) 63(1) Business Horizons 37 at 39.

[12] Ryan Calo "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51 UC Davis L Rev 399 at 404.

[13] TC "What are algorithms?" (30 August 2017) The Economist <www.economist.com>.

[14] Reyes and Ward, n 2, at 347.

[15] At 347.

[16] At 344.

Handcrafted algorithms, which are sometimes referred to as "expert systems" follow an if/then logic structure and require system designers to translate the knowledge of experts into a series of formal rules and structures that the algorithm-driven tool can then process.[17] A good example of a legal expert system is a software called TurboTax in the United States.[18] Imagine a tax law that says, for an income up to $20,000, an individual will be taxed at a marginal tax rate of 15.5%. A system designer will then translate this logic into an if/then computer rule such that if income < $20,000, then tax rate = 15.5%. While this is an oversimplified explanation of how the software operates, this explanation helps to illustrate the basic logic underlying handcrafted algorithms. Since these systems rely on experts' knowledge and clear rules that are well-established from the outset, handcrafted algorithms are only able to flourish in determinate domains where the legislative rules are laid out clearly ex ante and where facts are undisputed or uncontroversial.

Unlike the top-down approach used in handcrafted algorithms, machine learning algorithms follow a bottom-up approach where the computer algorithm organically determined its operating rules on its own.[19] The software "learns" by detecting patterns in large amounts of data, harnessing them to produce useful, intelligent-seeming decisions.[20] Surden gives a helpful example of a typical email spam filter to illustrate how machine learning works.[21] In the author's example, the machine learning system is firstly trained with multiple examples of spam emails and wanted emails. As more data is fed into the system, the system detects two indicia of spam – the words "free" and "Belarus". In this sense, the software used heuristics in automatically identifying which emails are spam emails. The example is not only helpful in understanding how machine learning algorithms operate, but also in understanding the limitations of such technology. Surden notes that although the software can detect a useful

---

[17] Richard E. Susskind "Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning" (1986) 49(2) MLR 168 as cited in Harry Surden "Artificial Intelligence and Law: An Overview" (2019) 35 Ga St UL Rev 1305 at 1316.

[18] Harry Surden "The Variable Determinacy Thesis" (2011) 12 Colum Sci & Tech L Rev 1 at 78.

[19] At 71 – 72.

[20] Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017-2018) 16 Duke L & Tech Rev 18 at 25.

[21] Surden, above n 17, at 1312.

pattern without a programmer having to explicitly program a specific set of rules, the phrase "learning" is only used as a metaphor for human learning.[22] This is because the software is unable to appreciate or understand the meaning behind the words. Machine learning algorithms also depend upon a large availability of data in order to function well.[23]

Knowing the capabilities and limitations of the aforementioned algorithms help us to understand how they would respond in different social contexts. The underlying idea here is algorithms should not be viewed as "computational instructions standing alone", but rather as algorithmic systems, requiring us to view algorithms as social technologies, imbued within a particular social context and interacting with actors in that context.[24] By considering the "contextual components of algorithmic systems" together with the systems' computational elements, Reyes and Wards submit that it will prevent us from underestimating or overestimating the potentials of these algorithmic-driven tools in our effort to improve the law.[25] Such an approach will help set realistic contextual goals. It allows us to compare algorithmic systems to the broken context of which it is currently a part of rather than measuring these systems against perfection. With this approach in mind, I will now give a brief overview of the sentencing law in New Zealand and its problems.

B The Limitations of the Sentencing System in New Zealand

Sentencing is known to be a notoriously complex and challenging area of criminal law despite it being a routine activity for judges. As McArdle J aptly puts it, "Anyone can try a case. That is as easy as falling off a log. The difficulty comes in knowing what to do with a man once he has been found guilty".[26]

Like any jurisdiction, the sentencing system in New Zealand is not free from flaws. In this part of my dissertation, I will first provide a brief outline of the sentencing system in New Zealand.

---

[22] At 1311.

[23] Dr Rajiv Desai "Artificial Intelligence (AI)" (23 March 2017) Dr Rajiv Desai An Educational Blog <https://drrajivdesaimd.com/>.

[24] Reyes and Ward, above n 2, at 344.

[25] At 344.

[26] Geoff Hall *Sentencing: 2007 Reforms in Context* (Lexis Nexis, Wellington, 2007) quoting McArdle J at 1.

Following that, I will highlight the flaws of the sentencing system in New Zealand with reference to four core judicial values that are relevant to the system. These core judicial values are transparency, impartiality, consistency and efficiency.

1 The sentencing system in New Zealand

The sentencing system in New Zealand affords a broad discretion to judges. Before the sentencing reforms in 2002, under the common law model, Parliament prescribed statutory maximum penalties and the types of sentences that were available to judges. The main statutory guidance was the maximum penalty of an offence which guides judges as to the upper limit in the worst class of cases. Subject to the particular maximum penalty, judges were afforded wide discretion to determine the appropriate sentence for an offender with three other sources of guidance and input that can influence sentencing decisions; appellate guidance wherein sometimes guideline judgments for specific offences are being produced, pre-sentence reports and the submissions made by the prosecution and the defence.

Then came the Sentencing Act of 2002 which specifies for the first time, the purposes of sentencing[27] and sets out a list of principles that judges must take into account when sentencing an offender.[28] The Sentencing Act also provides a non-exhausting list of aggravating and mitigating factors.[29] However, even with the new reforms, judges are still left with a wide discretion in determining sentences as no specific weighting was afforded for each purpose and principle outlined in the Sentencing Act. Although the Sentencing Act aims to provide a much more detailed legislative guidance, it still failed to fulfil the demands of sentencing law.

2 Problems with the current sentencing system

(a) Transparency

Considered fundamental to every democratic society, judicial transparency, often cited together with the principle of "open justice", is one of the most generally accepted judicial

[27] Section 7.

[28] Section 8.

[29] Section 9.

values.[30] Judicial transparency requires a "commitment to openness and candour" whereby the working and operations of the court are displayed for the public to see.[31] Transparency is closely related to explainability and accountability. The need for explainability in judicial decision-making is consistent with Bentham's ideas of giving reasons to enable an affected party to comprehend and criticise a decision.[32] Meanwhile, accountability is often understood as "the commitment to ensure that the values of independence and impartiality are appropriately deployed in the public interest, rather than the interest of the judges themselves".[33] Put simply, in the judicial context, transparency, accountability and explainability are vital in ensuring that individuals understand the reasons behind the decisions affecting them, leading to a better community understanding of legal domains hence increasing public confidence in judicial decision-making.

Given the subjective domain of sentencing law, it is understandable why judicial transparency remains an ongoing concern. Hall accurately describes this weakness of this judicial decision-making process:[34]

> Sentencing is not a rational mechanical process; it is a human process and subject to all the frailties of the human mind. A wide variety of factors, including the Judge's background, experience, social values, moral outlook, penal philosophy and views as to the merits or demerits of a particular penalty influence the sentencing decision.

Although the Sentencing Act provided a more transparent sentencing framework, it still gave limited guidance about the appropriate sentencing levels as it merely codified existing

---

[30] Monika Zalnieriute and Felicity Bell "Technology and Judicial Role" in Gabrielle Appleby and Andrew Lynch *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, Cambridge, 2021) 116 at 126.

[31] At 126.

[32] See Oren Ben-Dor "The institutionalisation of public opinion: Bentham's proposed constitutional role for jury and judges" (2007) 35 27 Legal Stud 216 as cited in Jessie Malcolm "Exploring the Enigma: Enhancing Digital Rights in the Age of Algorithms" (LLB (Hons) Dissertation, University of Otago, 2018) at 9.

[33] Richard Devlin and Adam Dodek "Regulating Judges: Challenges, Controversies and Choices" in Richard Devlin and Adam Dodek *Regulating Judges: Beyond Independence and Accountability* (Edward Elgar Publishing, Cheltenham, 2016) 1 at 9.

[34] Geoff Hall *Sentencing Law and Practice* (LexisNexis, Wellington, 2004) at [2.1].

case law. The multiplicity of purposes described in Section 7 would be helpful if relative weight of each purpose is specified and if the purposes all tended to point to the same direction. Unfortunately, this is not the case. Given such limited guidance in such a highly discretionary domain where complexity, uncertainty and time pressure are ongoing challenges, judges, due to their "bounded rationality", will rely on sentencing heuristics in search for "good enough" or "satisficing" decisions.[35] While some would argue that using heuristics in sentencing is reasonable due to the nature of the decision-making environment,[36] judicial transparency will remain a distant goal as it will be difficult to comprehend or reasonably understand the reasons behind a sentencing decision. However, judges' sentencing notes do help, to some extent, in improving transparency in judicial decision-making. I would argue, therefore, in assessing whether an algorithmic system would properly uphold the fundamental judicial value of transparency, it is wise to compare the algorithmic system to the transparency standard required of human judges in the current sentencing process. Once again, the goal here is comparative – we must measure the algorithmic sentencing system against the currently flawed sentencing process in New Zealand, instead of an ideal or perfect system.

(b) Impartiality

Impartiality, or fairness, defined as "the quality of not favouring one side or party more than another", is the cornerstone of a system of justice.[37] It mandates a judge, as the decision-maker to operate without the presence of bias or prejudice. Impartiality is vital both for individual determinations and retaining public confidence in the justice system. It is also a facet of equality or the dispensing of equitable justice, in that "like cases be treated

[35] Wayne Goodall "Sentencing Consistency in the New Zealand District Courts" (PhD Thesis, Victoria University of Wellington, 2014) at 35.

[36] Bettina von Helversen and Jorg Reiskamp "Predicting sentencing for low-level crimes: Comparing models of human judgment" (2009) 15 Journal of Experimental Psychology: Applied 375-395.

[37] Zalnieriute and Bell, above n 30, at 133.

alike".[38] In this sense, the notion of impartiality overlaps with the need for consistency in judicial decision-making.

Given the broad discretionary power afforded to judges, it is easy to understand why bias or prejudice might creep in at any stage of the criminal justice system. To illustrate the bias in the New Zealand sentencing system, one can look at the over-representation of Māori in prisons where Māori make up 52 percent of the prison population despite only making up 16 percent of New Zealand's total population.[39] Research also suggests that Māori convicted of assault are more likely to be imprisoned than Europeans, despite both being found guilty of the same crime.[40] These alarming statistics tend to suggest that the New Zealand criminal justice system is one that primarily targets Māori.

Impartiality is critical in considering the influence of an algorithmic system on the judicial role. While protected variables such as race and gender can be excluded from the system, discrimination by proxy is a significant concern for algorithmic sentencing systems.[41] On the other hand, stripping bias from the data set, could make the data less useful as it slices up the data into such small pieces that hardly anything is left.[42] This issue will be explained further in chapter II. Looking at New Zealand's criminal justice system that disproportionately affected Māori, designing an algorithmic system that does not replicate bias is essential. On that note, it is vital therefore, when designing an algorithmic sentencing system, to consider the perspectives of individuals who are likely to be affected by algorithmic decisions and those likely to be under-represented in construction and training

---

[38] John Rawls *A Theory of Justice* (Oxford University Press, Oxford,1999) at 237; and HLA Hart "Positivism and the Separation of Laws and Morals" (1958) 71(4) Harv L Rev 593 at 623-624 as cited in Monika Zalnieriute, Lyria Bennett Moses and George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82(3) MLR 425 at 431.

[39] Jendy Harper "Why Does NZ Imprison So Many Maori" *Newsroom* (online ed, New Zealand, 29 August 2020).

[40] Jordan Bond "Maori Imprisoned at Twice Rate of Europeans for the Same Crime" *NZ Herald* (online ed, New Zealand, 14 September 2016).

[41] Anupam Datta, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, Shayak Sen "Proxy Discrimination in Data-Driven Systems: Theory and Experiments with Machine Learnt Programs" (2017) Cornell University at 3.

[42] Will Knight "AI is Biased. Here's How Scientists Are Trying to Fix It" (19 December 2019) Wired <https://www.wired.com/>.

of the algorithmic systems.[43] This is likely to include those in lower socio-economic classes, Māori and Pacific Island populations.[44] Such an approach is consistent with one of the commitments outlined in New Zealand's Algorithm Charter which requires government agencies to include Te Ao Māori perspectives in the development and use of algorithms.[45] I will expand on this further in chapter III.

However, it is wise to note that compounded racial bias is a structural problem in criminal justice. It is an illustration of a broader systemic issue that continues to discriminate and oppress Māori in their homeland regardless of whether we utilise structured algorithms or human judgment at a given stage of the process.[46] This is not to understate the importance of mitigating bias in these algorithmic systems but rather, to consider that algorithmic systems may not be the panacea for the sources of discriminatory outcomes that occur independently of sentencing.

(c) Consistency

Consistently applying judicial decisions is a fundamental element in upholding the rule of law. Sentencing consistency, or the extent to which like cases are treated alike, has been given statutory recognition in New Zealand under the Sentencing Act.[47] Consistency fosters transparency and predictability in sentencing practices, maintains public confidence in the judicial decision-making process and promotes legitimacy of the criminal justice system itself.[48] The importance of sentence parity has been acknowledged by the New Zealand Court of Appeal, which held that inconsistency in sentencing decisions "can result in

---

[43] Gavaghan, Knott, Maclaurin, Zerilli and Liddicoat, above n 4, at 77.

[44] At 77.

[45] Stats NZ "Algorithm charter for Aotearoa New Zealand" (20 November 2020) data.gov.nz <https://data.govt.nz>.

[46] Harper, above n 39.

[47] Section 8(e).

[48] Julian V. Roberts and Mojca M Plesnicar "Sentencing, Legitimacy and Public Opinion" in Gorazd Mesko and Justice Tankebe *Trust and Legitimacy in Criminal Justice: European Perspectives* (Springer, Cham, 2015) 33 at 47.

injustice to an accused person and may raise doubts about the evenhanded administration of justice".[49]

New Zealand judiciary enjoys considerable discretion in comparison to some international jurisdictions like the United States. While guideline judgments and the Sentencing Act exist to guide judicial discretion, New Zealand judges still maintain considerable discretion to determine the appropriate sentencing levels. There is limited research done in New Zealand regarding sentencing disparity, but existing studies show an inconsistency in how sentencing judges determine the sentence to impose.[50] The Federal Sentencing Guideline and the Sentencing Information System are two examples of past efforts introduced in other jurisdictions to achieve consistency in sentencing.

The United States have introduced a "Federal Sentencing Guideline" or "Guidelines" in an effort to curb judicial discretion in exchange with promoting consistency, transparency and predictability into the sentencing regime.[51] Based on both the seriousness of the crime and the offender's criminal history, these sentencing grids are used to calculate the applicable sentence.[52] Until today, the sentencing guidelines remain controversial in the United States with some claiming the guidelines are "too complex, inflexible, and severe".[53]

Another method in curbing sentencing disparity is the Sentencing Information Systems (SIS) – sometimes referred to as Decision Support Systems and sentencing databases. An example of SIS would be the one in New South Wales which gives judges quick access to a database consisting of legal, factual and statistical data about predominant sentencing

---

[49] *R v Morris* [1991] 3 NZLR 641 (CA) at 645.

[50] Samantha Jeffries, Garth Fletcher and Greg Newbold "Pathways to Sex-Based Differentiation in Criminal Court Sentencing" (2006) 41 Criminology 329 at 347; and Wayne Goodall and Russil Durrant "Regional Variation in Sentencing: The Incarceration of Aggravated Drink Drivers in the New Zealand District Courts" (2013) 46 ANZJ Crim 422 at 441.

[51] Alexis Lee Watts "In Depth: Sentencing Guideline Grids" (11 January 2018) University of Minnesota <https://twin-cities.umn.edu/>.

[52] Watts, above n 51.

[53] George Coppolo "Criticisms of Federal Sentencing Guidelines" (16 July 1994) The Connecticut General Assembly <https://www.cga.ct.gov/PS94/rpt/olr/htm/94-R-0686.htm>.

patterns enabling the judge to discover the range of penalties imposed in the past for similar offences.[54] While SIS provides no guidance about what the appropriate sentence should be, the aim was that by displaying statistical information about past sentences, a judge, wishing to pursue consistency, would choose a sentence within the statistical average.[55] Therefore, SIS operates with the idea of striving for consistency while retaining judicial independence of the sentencing process.[56] Few databases, however, are still operating and explanations for their failure are partly due to a lack of institutional support and judicial apathy.[57] These are also significant concerns for the successful implementation of the proposed algorithmic sentencing system I envision to be implemented in New Zealand.

The two methods I have outlined – the Federal Sentencing Guideline and SIS – illustrate the difficulties in achieving consistency in sentencing. Importantly, there exists a perennial conflict in finding a suitable balance between achieving consistency in sentencing practices and upholding the principle of individualised justice.[58] How to resolve this conflict and finding an appropriate equilibrium is a legal challenge faced by jurisdictions worldwide. As such, algorithmic decisions are often criticised for not being sufficiently individualised. I will address this concern in the next part of this chapter.

(d) Efficiency

Efficiency is an emerging judicial value that is gaining momentum in the legal system. Usually considered a subset of accountability, judicial efficiency has now been recognised as a stand-alone variable.[59] Devlin and Dodek define efficiency as "the aspiration that social

---

[54] Katja Franko Aas *Sentencing in the Age of Information: from Faust to Macintosh* (The GlassHouse Press, London, 2005) at 33.

[55] At 32.

[56] At 32.

[57] Cyrus Tata "The Application of Judicial Intelligence and "Rules" to Systems Supporting Discretionary Judicial Decision-Making" in G. Sartor and L. Karl Branting *Introduction: Judicial Applications of Artificial Intelligence* (Kluwer Academic Publishers, Dordrecht, 1998) 203 at 211.

[58] Sarah Krasnostein and Arie Freiberg "Pursuing Consistency in an Individualist Sentencing Framework: If You Know Where You're Going, How Do You Know When You've Got There?" (2013) 76 LCP 265 at 265.

[59] Zalnieriute and Bell, above n 30, at 126.

and personal investments in the judiciary and judicial processes are cost-effective".[60] Efficiency involves not only the greatest utilisation of the scarce resources of the judiciary but also the timely delivery of judgments.[61] Perhaps the maxim "justice delayed is justice denied" best encapsulates the notion that efficiency is of fundamental importance to access to justice.

New Zealand, like most jurisdictions, suffers from a backlogged court system that has been further exacerbated by COVID-19.[62] The goal of efficiency is usually brought forward to introduce technology to a new setting, especially in the legal domain. In my proposal to introduce an algorithmic sentencing system into New Zealand, my main argument is that such a system will improve the efficiency of our justice system. An algorithmic sentencing system would be able to analyse a large amount of past sentencing decisions and generate a sentence prediction quicker than a human judge would, thus making the process more efficient and preventing unnecessary delays in the justice system.

However, it is essential to note here that while the maxim "justice delayed is justice denied" is attractive to some extent, it risks oversimplifying the situation. Focusing on quantitative measures of efficiency instead of qualitative explanatory factors tends to overlook other fundamental judicial values that the judiciary must safeguard, including an independent, fair and impartial judicial process.[63] Explaining further, Bathurst says that by emphasising efficiency too much in order to cut costs and produce quicker judgments, it can risk undermining the quality and thoroughness of the judicial decision-making process itself and create wider legal uncertainty and instability.[64] Hence, the push to efficiency fails to

---

[60] Devlin and Dodek, above n 33, at 9.

[61] Gabrielle Appleby and Heather Roberts "The Chief Justice: Under Relational and Institutional Pressure" in Gabrielle Appleby and Andrew Lynch *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, Cambridge, 2021) 50 at 69.

[62] Emile Donovan "Delayed Justice – Our Courts Under Pressure" *Newsroom* (online ed, New Zealand, 28 July 2020).

[63] The Honourable TF Bathurst AC "Who judges the judges, and how should they be judged?' (2019) 14 TJR 19 at 37.

[64] At 36.

recognise that "justice rushed" is as much a failure as "justice delayed", as the parties involved in a sentencing decision as well as the wider community deserve properly considered judgment that may be appropriately cost and time-consuming.[65]

On that note, the question of whether we should implement an algorithmic system to the sentencing process must be approached, bearing in mind that efficiency in the administration of justice must be balanced with other fundamental rights and principles.

C How Might an Algorithmic Sentencing System Improve the Sentencing System?

My primary argument for introducing an algorithmic sentencing system in New Zealand rests mainly on the goal of efficiency. While consistency is often cited as one of the main reasons for implementing an algorithmic sentencing system, there are concerns that using algorithmic driven tools to achieve consistency might not be desirable.

1 Improving efficiency: a balanced concept

In largely uncontroversial legal areas, algorithms are used in document retrievals and legal research, minimising the use of paper and allowing quick access to court's transcripts which are beneficial in increasing efficiency.[66] However, technology proponents also rely on the efficiency argument to support introducing algorithms in more controversial areas of the law such as the automation of small claims or minor offences and risk assessment tools.[67] A further example where technology is used in a much more complex legal domain is in the legal context concerning social security where government departments are using algorithms designed to determine whether an individual is eligible for a wide range of government benefits.[68] Algorithm systems are also used to determine whether an individual is eligible to come to and

---

[65] JJ Spigelman, "The quality dimension of judicial administration" (1999) 4(3) TJR 179 at 184.

[66] Zalnieriute and Bell, above n 30, at 138.

[67] Nikolaos Aletras, Tsarapatsanis Dimitrios, Preoţiuc-Pietro Daniel, and Lampos Vasileios "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective" (2016) 2 PeerJournal of Computer Science 93; and Felicity Bell "Family Law, Access to Justice, and Automation" (2019) 19 Macquarie LJ 103.

[68] Department of Human Services, Australian Government "Ticking All the Right Boxes" (4 September 2017) Australian Government Services Australia < https://www.servicesaustralia.gov.au/>.

stay in the country.[69] Such high-stake decisions involving voluminous and complicated considerations, are certainly no less complex than sentencing law.

Given the impressive and rapid development of machine learning algorithms that are increasingly gaining attention nowadays and more promisingly in the future, it seems tenable to design a reasonably sophisticated algorithmic sentencing system that harmonises an algorithmic process with a judicial decision-making process that requires considerable human judgment and involvement. I would argue that such a reasonably sophisticated algorithmic sentencing system while able to generate quicker sentence predictions, would also be able to generate more accurate and granular predictions, thus improving efficiency while maintaining the quality of the sentencing outcome. A backlogged court system, like the one faced by the justice system in New Zealand, would benefit from this.

It is also important to stress here that an algorithmic sentencing system must be accurate as an error with a widely applied sentencing algorithm could rapidly scale up to disadvantage a large number of cases, as compared to a human error. The Australian "robo-debt" controversy is a good example.[70] Robo-debt was an automated decision system used by the Australian government to determine if beneficiaries had been overpaid. Mistakes in its application meant that the Australian government has had to pay back over one billion dollars. Therefore, it is vital that an algorithmic sentencing system is accurate to prevent a similar scenario from occurring, especially given the high-stake considerations and cardinal interests involved in sentencing an individual. This is where a pre-deployment accuracy checking process will be useful as it will ensure the quality of the sentencing predictions in the pursuit of efficiency. I will explain more on this in chapter IV.

However, the question arises of what would be considered as an accurate prediction for a sentencing decision. Would an algorithmic sentencing system be accurate if it generated

---

[69] Department of Immigration and Border Protection, Australian Government "Applying online or paper" Australian Government Department of Home Affairs < https://immi.homeaffairs.gov.au/>.

[70] Luke Henriques-Gomes "Robodebt class action: Coalition agrees to pay $1.2bn to settle lawsuit" *The Guardian* (online ed, United Kingdom, 16 Nov 2020).

equivalent results to human judges? Given the flaws in human decision-making, this might not be desirable. However, if we want it to be better than human judges, how do we measure that? There is plenty of research done around this area and the question is still left unanswered. For the purpose of my dissertation, I will consider an accurate prediction as one that generates similar results to human judges.

2 Promoting consistency via algorithmic systems: a misleading goal?

In his recent work, Chiao contends that a sentencing algorithm based on a machine learning approach can minimise the unjustifiable disparity in imposed sentences.[71] His claim is based on a thought experiment and can be summarised with reference to the two figures below:[72]

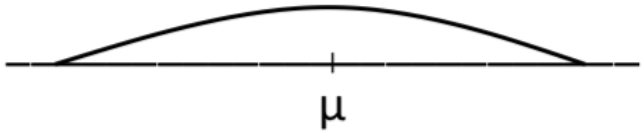

Figure 1: Sentence distribution without the support of a sentencing algorithm

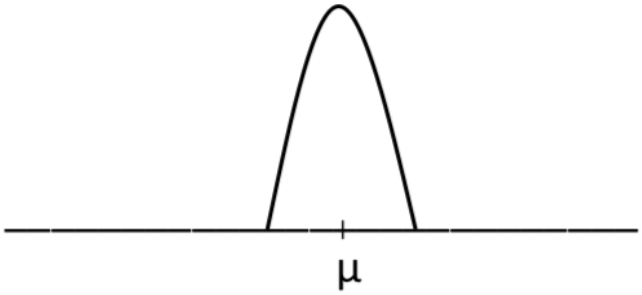

Figure 2: Sentence distribution with the support of a sentencing algorithm

[71] Chiao, above n 3, at 246.

[72] At 247.

According to Chiao, a sentencing algorithm can reduce discrepancies in a system of discretionary sentencing by making the sentence distribution somewhat narrower i.e., reducing variance by encouraging a transition of sentencing distribution illustrated in Figure 1 to the one shown in Figure 2. In Figure 1, judges are making sentencing decisions for similar offences without the assistance of an algorithmic sentencing system. Given that different judges will give different weight on relevant sentencing factors, purposes and principles, it is highly likely that there will be a certain degree of severity relating to interjudge variation in the sentencing outcomes. However, suppose the machine learning algorithm is used by human judges in sentencing offenders of similar crimes. In that case, we could expect that the sentences will be more narrowly distributed around the mean as shown in Figure 2, albeit with some variation. By having a sense of what they and their peers have deemed as proportionate in other similar cases, the sentencing predictions act as a psychological anchor that the parties can draw upon as common knowledge.[73]

Despite being an empirical claim, Chiao's argument that the aforementioned procedure would reduce sentencing disparity seems to be reasonable. For instance, one of the explanations for inconsistency in sentencing decisions is that judges are not fully aware of previous sentences that their colleagues have passed.[74] If this is correct, it seems plausible to say that the algorithmic sentencing system does, in fact, lead to convergence in interjudge sentencing decisions. However, there is a concern that achieving consistency in such a way might not be desirable. The issue here is to what extent do we want judges to consider other factors to ensure sufficient individualisation? This is also a question I pose to the Sabah and Sarawak courts with regard to AICOS. When asked about the numerous factors that judges would consider in sentencing decisions, the representatives of the High Court in Sabah and Sarawak said that regarding the drug-related offence, they have identified beforehand the factors that are most pleaded for that offence.[75] Based on such information, judges will then input the

[73] Kiel Brennan-Marquez and Vincent Chiao "Algorithmic Decision-Making When Humans Disagree on Ends" (2021) 24(3) New Criminal Law Review 275 at 299.

[74] Jesper Ryberg "Sentencing Disparity and Artificial Intelligence" The Journal of Value Inquiry 1 at 4.

[75] AI Committee Members of Sabah and Sarawak, above n 6.

aforementioned relevant factors into the system. Judges are then allowed to depart from the sentencing prediction if they consider that there are other factors and parameters which are not available in the system.[76] The court hopes that, with time, the system will improve to be able to consider a larger number of factors.

Such an approach is consistent with my modest case for an incremental establishment of an algorithmic sentencing system in New Zealand. What could be done here is, once we have identified the relevant factors often considered for a particular offence, potentially by carrying out a survey among members of the judiciary i.e., judges, we could achieve the outcome in Figure 2, but with all the relevant factors included. If the judge identifies any factor that is not included that should play an important part in predicting the sentencing outcome, the judge may depart from the algorithm system's recommendation. In this sense, I submit that an algorithmic sentencing system would contribute to achieving consistency in the sentencing process in New Zealand by introducing a publicly known and predictable baseline but based on the most pleaded factors for that offence. The benefits derived from such a scenario could potentially extend beyond the courtroom where lawyers and defendants could use the algorithm sentencing system to calculate the risk of whether or not to go to trial.[77]

D Conclusion

Drawing on the framework developed by Reyes and Ward, I have shown that within the social context of sentencing, judicial values such as transparency, impartiality, consistency and efficiency are of fundamental importance to maintain a legitimate criminal justice system that upholds the rule of law and maintains public confidence. However, these values do not operate independently and sometime do come in conflict with one another.

What I hoped to illustrate in this chapter is that there is a gap in the current sentencing system in New Zealand in terms of the aforementioned judicial values, and such a gap is also partly attributed to the multifarious factors beyond sentencing law. When introducing a new solution to change the status quo, the issue is comparative – how can algorithmic systems do better

[76] AI Committee Members of Sabah and Sarawak, above n 6.

[77] Chiao, above n 3, at 256.

than the current flawed system and unfortunately, that is currently a low bar. An algorithmic system is not the panacea to everything that is wrong with the criminal justice system, especially issues that are imbued within the system itself, but managing realistic expectations, I argue, would be a partial step in the right direction to improve the law. This is the appropriate mindset in approaching the issue of whether an algorithmic system would be suitable in the domain of criminal sentencing.

I have also illustrated how an algorithmic sentencing would help to improve efficiency in the justice system and to some extent, consistency. In the next chapter, I will illustrate what an algorithmic system for criminal sentencing in New Zealand could look like, borrowing from Vincent Chiao's recent work. I will also make some reference to the AI sentencing model in Malaysia. I will also point out some concerns in terms of transparency, impartiality and judicial ambivalence in introducing an algorithmic sentencing system in New Zealand.

CHAPTER II: AN ALGORITHMIC SENTENCING SYSTEM FOR NEW ZEALAND

A What Could an Algorithmic Sentencing System in New Zealand Look Like?

In my proposal for an algorithmic sentencing system in New Zealand, I will refer to the work of Chiao, who recently presented and defended the use of a machine learning algorithm to be used in supporting judges in the sentencing process.[78] Chiao's overall idea is generally straightforward.

In the first step of Chiao's recommendation, what is required is building a rich dataset consisting of previous sentencing decisions in the relevant jurisdiction. The next step would require a judge to provide the algorithm with input relating to the relevant sentencing factors of a present case. As mentioned previously, since many aggravating and mitigating factors could influence a sentencing decision, a survey could be done beforehand among members of the judiciary to identify the most relevant factors that are often used for that particular offence. Based on the result of the survey, the judge will then insert such factors into the system. The system will then analyse this information and deliver output in the form of a sentence prediction as well as "sentences within a standard deviation from the average", based on the judiciary's own sense of what have previously constituted appropriate punishments.[79] The recommendation made by the algorithmic system would not be binding on judges as they can depart from the algorithm's recommendation where the case before them is highly unusual in some significant respects i.e. where the court considers other factors and parameters which are not available in the system. In such cases, judges are required to give reasons for such a departure by explaining in what respect the case before the judge is unique and unusual. This requirement would prevent the sentencing process from becoming "an exercise in rubber-stamping"[80] as the algorithm's predictions serve merely "to inform judges of what has been

---

[78] Chiao, above n 3, at 245.

[79] At 240.

[80] At 246.

deemed proportionate in cases exhibiting a similar constellation of relevant factors".[81] Chiao claimed that the algorithm system would assist judges in providing "a particularised snapshot of the central tendency of how they and their colleagues have been treating similar cases".[82]

While handcrafted algorithms would be more transparent and easier for human comprehension, a machine learning algorithm promises more refined and accurate predictions.[83] In supporting the use of machine learning algorithms rather than handcrafted algorithms based on traditional regression techniques, Chiao argues that a machine learning algorithm can discover correlations on its own, thus able to revise its prediction in light of new judgments dispensing the need for human intervention to continually update the system with judicial opinion.[84] With a machine learning approach, Chiao claims that, provided that we have a rich set of input data, the algorithm does not need to be encoded with any specific sentencing theories like proportionality, as it will learn "correlations of its own between input features and outcomes".[85] This is based on the assumption that we have a rich data set of sentencing decisions which highlight the range of punishment that would be proportionate.

It is essential to clarify that the algorithm system that Chiao envisioned is not to model the kind of reasoning that human judges experienced in deciding a sentence, as Chiao acknowledged that such a process would involve "a rich moral tapestry of actions, intentions, emotions, harms and relationships",[86] but rather "to provide a reliable and accurate prediction of what a typical sentencing judge in the relevant jurisdiction would regard as proportionate on a given set of facts".[87] In essence, the algorithm system does not seek to replicate the moral content of the sentencing process. Instead , it seeks "to predict outcomes related to how

[81] At 243.

[82] At 246.

[83] At 250.

[84] At 245.

[85] At 245.

[86] At 246.

[87] At 246 – 247.

human judges apply that concept in practice".[88] He analogises this with algorithms that have been used to predict people's taste in music or film.[89]

Chiao's recommendation for a sentencing algorithm differs from the current New Zealand approach to sentencing in two ways. Firstly, the algorithm bypasses the two steps of "the *Taueki* method" and goes straight to predicting the appropriate sentence. In *R v Taueki*, the Court of Appeal established "the *Taueki* method" which is essentially a two-step approach to sentencing that applies across all types of cases.[90] The first step involves the selection of a starting point appropriate to the seriousness of the offending followed by the second step which requires judges to adjust the starting point upwards or downwards to reflect factors personal to the offender.[91] In this sense, the kind of uniformity that is sought by requiring judges to apply starting points is instead built into the correlations that the algorithm learns from the data fed into it. The court can then consider any case-specific aggravating and mitigating factors that the algorithmic system fails to take into account and then adjust the sentencing prediction accordingly. Secondly, the machine learning algorithm's ability to learn from experience will allow it to update its prediction in light of the success or failure of its prior predictions. Concerning the latter, in the event where judges decide to depart significantly from the algorithm's predictions, such a departure would provide a basis for the algorithm to update its future predictions, thus making the system more accurate and up to date with the progress in judicial opinion.[92]

Chiao's recommendation, while attractive, has its limitations. His proposal for an algorithmic sentencing system lies on the assumption that 1) There is a sufficiently rich data set of sentencing decisions on a particular offence as machine learning algorithms' ability to function effectively depend on massive availability of data and 2) The machine learning algorithms' ability to map distinct permutations of factors that the judiciary has so far acknowledged to be morally relevant to punishments, which would be a complex matter given the large number of

---

[88] At 247.

[89] At 247.

[90] *R v Taueki* [2005] 3 NZLR 372 (CA).

[91] At 44.

[92] Chiao, above n 3, at 251.

such factors. However, for the purpose of the dissertation, I will consider these assumptions true as they are not entirely impossible to achieve, albeit being a complex and challenging task. This is acknowledged by Chiao in his work.[93] Indeed, the Courts of Sabah and Sarawak took about six months for data identification, data collection, machine learning and software development, consultation with court users and stakeholders, execution as well as monitoring and maintenance to ensure the successful implementation of AICOS. [94]

Consistent with my modest and incremental approach for an algorithmic sentencing system in New Zealand, I envision the system to be trialed at the sentence indication stage. A sentence indication is a statement by the court that provides an individual with an idea of the type or quantum (amount or length) of the sentence they would be likely to receive if they were to plead guilty at this stage of their proceeding.[95] During the early phases of implementation, the system could be applicable for common, low-level offences first such as traffic-related offences where the most pleaded factors can be easily identified. Then, suppose the pilot project proves to be successful and this is measured by the level of accuracy exhibited by the system, I see the system being operated in New Zealand's sentencing process for more severe offences in two scenarios: where the accused has pleaded guilty, or when the court finds the accused guilty after full trial. As mentioned in chapter I, I consider an algorithmic sentencing system accurate if it generates similar sentence predictions to human judges.

Now, I will proceed to evaluate the limitations of an algorithmic sentencing system in New Zealand.

B The Limitations of an Algorithmic Sentencing System

1 Open justice – transparency, accountability and explainability

One of the main reasons for the resistance against a machine learning approach to sentencing is that such systems can be opaque. This would negatively affect judicial transparency as it will

---

[93] At 241.

[94] AI Committee Members of Sabah and Sarawak, above n 6.

[95] Ministry of Justice "Sentence indication" < https://www.justice.govt.nz/>.

make it difficult for an individual who has been subjected to an algorithmic decision to review or challenge the decision made with the assistance of an algorithmic system.[96] In decisions involving high-stake consequences like criminal sentencing, it is of the greatest importance that those affected by the decision can meaningfully understand the underlying reasons behind such decision, critically analyse whether the imposed sentence is both fair and reasonable and hold those responsible for such decisions to account. Therefore, in this respect, the usual question posed by opponents of sentencing algorithms is that if traditional sentencing is more transparent than a sentencing algorithm, wouldn't human sentencing be more morally advantageous for the criminal justice system?

To properly assess this claim, it is paramount to discuss how exactly transparency in an algorithmic sentencing system works. As a starting point, it is helpful to think of judicial transparency for an algorithmic system as a function of two factors: access and complexity.[97] Despite there being no precise mathematical definition, it is widely accepted that a transparent algorithmic system is one where an individual who is subjected to such a system – especially where their life is significantly affected by the outcome generated by the algorithm – can understand fully and meaningfully how the algorithm system came to its prediction.[98]

In terms of accessibility, access to information about the algorithmic system in terms of its source code or a mathematic representation of the model is the first important step to algorithmic transparency. The difficulties surrounding access to algorithms are mainly due to commercial sensitivity. This is due to the fact that private companies are usually the drivers behind such algorithmic systems and clearly have an interest to keep information relating to

---

[96] Frank Pasquele *The Black Box Society: The Secrets Algorithms That Control Money and Information* (Harvard University Press, Cambridge, 2015) as cited in Bruno Lepri, Nuria Oliver, Emmanuel Letouze, Alex Pentland and Patrick Vinck "Fair, Transparent and Accountable Algorithmic Decision-Making Process" (2017) 31(4) Philos Technol 612 at 619.

[97] Frej Klem Thomsen "Iudicium ex Machinae – The Ethical Challenges of Automated Decision-Making in Criminal Sentencing" (October 2020 draft, Danish Institute for Human Rights, Copenhagen, 2020) 1 at 8.

[98] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi "A Survey of Methods for Explaining Black Box Models" (2018) 51(5) ACM Comput Surv 1 *at 5*.

the algorithm's model and source code confidential.[99] Indeed, this is also the case with AICOS.[100] The controversial case of *State v Loomis* is a perfect illustration to demonstrate the danger of withholding information in relation to an algorithm system based on trade law restrictions.[101] In that case, Eric Loomis challenged the State of Wisconsin's use of closed-source risk assessment tool that the sentencing judge relied on to sentence him to six years' imprisonment. Loomis contended that the software violated his right to due process because it prevented him from challenging the scientific validity and accuracy of such data. He demanded more information concerning the algorithm, but the court could not share it as such information was trade secrets. This decision illustrates the harms of algorithmic decision-making when it comes to accessibility.

Nevertheless, there is no reason why an algorithmic sentencing system should not be made fully available to the public. One way to do this is by requiring government agencies to develop the algorithms in-house or by a contract that allows for such transparency.[102] This is also what the judiciary in Sabah and Sarawak is planning to do in the future which is to have the judiciary owning the AI model to ensure that defendants could see how the system works.[103] Such an approach, Briony Blackmore argues, has two important benefits. Firstly, courts are not left to the mercy of private companies who can choose not to disclose information relating to the algorithm decision relying on trade law restrictions. Secondly, the algorithm system can be designed with a particular aim and context in mind.[104]

Even if information relating to the algorithmic system is legally accessible, it is important that such information is not too complex for those subjected to the algorithm's decision to understand how the system properly works. However, when it comes to machine learning

---

[99] Taylor R Moore *Trade Secrets and Algorithms as Barriers to Social Justice* (Centre for Democracy & Technology, Washington, 2017) at 5.

[100] AI Committee Members of Sabah and Sarawak, above n 6.

[101] *State v Loomis* 881 NW 2d 749 (Wis 2016).

[102] Briony Blackmore "Developing Transparency Requirements for the Operation of Criminal Justice Algorithms in New Zealand" (PhD Thesis, University of Otago, 2019) at 95.

[103] AI Committee Members of Sabah and Sarawak, above n 6.

[104] At 67.

algorithms, there are technical barriers to developers and operators to explain how the algorithm system work. For instance, the Harm Assessment Risk Tool (HART), which was developed in Durham to reduce future harm and risks of recidivism, has 4.2 million decision points in which its developers are not able to explain how their software works.[105] This is also a potential problem for the algorithmic sentencing system mentioned earlier as the system, in order to come up with a very complex decision process, would be using datasets that would presumably involve a great many variables, weightings, and decision points. On that note, it is also important to evaluate what level of transparency in terms of explainability is required for an individual to understand the algorithmic sentencing decision as "full transparency" would potentially lead to privacy concerns and people "gaming the system".[106] Delving into complicated technicalities to explain the inner workings of the system also would not be beneficial.

It is worth remembering that even the human sentencing process is far from transparent. As pointed out by several experts, even in the realm of human decision-making, it is difficult to observe directly how an individual came to a decision and even in the event where reasons are sought for such a decision, there is an unfortunate tendency for people to give inaccurate reasons.[107] Indeed, it is naive to think that the reasons that people disclose always overlap with the actual reasons behind their decisions; we can even fool ourselves about why we made such decisions.[108] However, that does not mean that efforts of algorithmic transparency should be abandoned. What I hoped to clarify here is that when it comes to improving algorithmic transparency in the sentencing domain, explanations for an algorithmic decision should be pitched at a practical level, one that is measured against the current standard required of human judges.

---

[105] 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) I & CTL 223 at 234.

[106] David J. *Gunkel Gaming the System: Deconstructing Video Games, Game Studies, and Virtual Worlds* (Indiana University Press, Bloomington, 2018) at 1.

[107] Vincent Chiao "Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice" (2019) 15(2) Int JLC 126 at 136.

[108] Daniel Kahneman *Thinking, Fast and Slow* (1st ed, Daniel Kahneman, New York, 2017) at 195.

2 Impartiality

Algorithmic systems are often touted as being impartial as compared to humans in the sense that they will not exhibit bias or prejudice in the decision-making process. I will now show how such a claim is flawed. In particular, I will focus on how an algorithmic sentencing system could be biased against vulnerable minorities hence exacerbating and reproducing existing inequalities.

Algorithmic systems can be biased in two ways – it can replicate bias in training data and during the software development.[109] In the data training stage, an algorithmic system that is trained on historical data that is inherently prejudicial against a particular group – mainly attributed to wider, systemic issues – will, as a consequence, perpetuates such existing bias. A famous phrase to explain the aforementioned phenomenon is 'garbage-in, garbage-out', whereby an algorithmic system is only as good as its training data. This has been clearly shown in the vast literature on risk assessment tools used in the criminal justice system.[110] Outside of the sentencing domain, one example where an algorithmic tool perpetuated existing bias was Amazon's recruiting tool, which relied on historical recruitment data that consistently downgraded female candidates during its data training stage.[111] Even when variables such as race or gender are removed from the system, proxy variables such as postcode, which is often said to operate as a proxy for race, may result in discrimination. In essence, what the aforementioned examples demonstrate is that "we cannot expect an AI algorithm that has been trained on data that comes from society to be better than society – unless we have explicitly designed it to be."[112] Even at the problem structuring stage, where designers of an algorithmic sentencing system need to decide what they want to algorithmic system to achieve,

---

[109] Thomsen, above n 97, at 10.

[110] Bernard E. Harcourt *Against Prediction: Profiling, Policing, and Punishing in An Actuarial Age* (University of Chicago Press, Chicago, 2006); Kelly Hannah-Moffat "Actuarial Sentencing: An 'Unsettled' Proposition" (2012) 30(2) JQ 270; and Sonja B. Starr "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination." 66 *Stan L Rev* 803.

[111] James Vincent, "Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Bias Against Women" (10 October 2018) The Verge <https://www.theverge.com/>.

[112] Bernard Marr "Artificial Intelligence Has a Problem with Bias, Here's How to Tackle It" *Forbes* (online ed, United States of America, 19 January 2019).

the parameters chosen by the developers may reflect their intentions or subconscious biases.[113] For instance, Beauty. AI was a software developed to have technology evaluate beauty objectively. However, when designing the algorithm, the developers unconsciously reinforced their own beauty standards when creating the algorithm.[114] As a result, the winners of the beauty contest were mostly "white".

It is indeed difficult to strip human bias and prejudice from the algorithm themselves.[115] However, bias is not always a bad thing. Bias may be desirable; for instance, we may want technologies used to detect cancerous tumours to be somewhat biased on the side of caution, tagging scans as risky even where there is uncertainty. Bias may also be an inbuilt policy decision in the public sector. An example of this is HART, where an offender is more likely to be categorised as medium or high-risk as the system itself is created specifically to be on the side of caution in terms of public safety.[116] However, the question of who is most suitable to make such a decision is contentious and this is a concern that I will also address in chapter IV.

3 Judicial thinking under attack?

When it comes to algorithmic sentencing systems, judicial ambivalence is a central issue. Whether judicial sentencers would endorse an algorithmic sentencing system or not is of paramount importance as legitimation would be vital to the longevity of any judicial decision support system.

---

[113] Karen Hao "This is How Bias in AI Really Happens – And Why It's So Hard to Fix" *MIT Technology Review* (online ed, Cambridge, 14 February 2019).

[114] Noel Duan "When Beauty is in the Eye of the (Robo)Beholder" (20 March 2017) Arstechnica <https://arstechnica.com>.

[115] Will Knight "The Dark Secret at the Heart of AI No One Really Knows How the Most Advanced Algorithms Do What They Do That Could Be a Problem" *MIT Technology Review* (online ed, Cambridge, 11 April 2017).

[116] Chris Baranjuk "Durham Police AI to help with custody decisions" *BBC* (online ed, London,10 May 2017) as cited in Malcolm, above n 32 at 12.

As described by Hall, sentencing is not a rational mechanical process but rather a human process.[117] However, by reducing the sentencing process to an algorithmic method, human variables arguably, have to a significant extent, disappeared from the sentencing process. This is because computerised representations of sentencing consistency tend to decontextualise and dehumanise sentencing.[118] Tata described this problem as a "loss of narrative".[119] Under the traditional sentencing systems, human judges, when deciding the appropriate sentence for an offender, will meaningfully assess the character of the individuals and the material realities of their lives.[120] In this sense, they are able to take into account variables that even the most careful programming could not account for.[121] Such judicial insightfulness, as a result of years of experience, local knowledge and professional values is arguably, however, increasingly restricted by algorithmic systems. This is because such systems tend to decontextualise the affected party, making the affected party a "collection of characteristics" rather than a whole person with a sound identity to make information more portable and easily consumable. [122]

I have asked the representatives of the Sabah and Sarawak court about judicial satisfaction with regard to AICOS.[123] Although no survey has been conducted concerning judicial satisfaction, statistics have shown that there has been prevalent use of AICOS, with more than 90% of cases using the system when sentencing an offender.[124] At the same time, feedback is consistently being asked from the Magistrates as to what the algorithmic sentencing system can improve on. Magistrates are also consistently reminded to exercise their human intelligence

---

[117] Hall, above n 34, at [2.1].

[118] Jacqueline Tombs "Telling sentencing stories" in Pat Carlen *Imaginary Penalties* (1st ed, Willan Publishing, Cullompton, 2008) 84; and Katja Franko Aas "From narrative to database: Technological change and penal culture" (2004) 6(4) Punishment & Society 379 as cited in Cyrus Tata "The Rise of Technology and the Demise of the Sentencing Professions?" in Dave Cowan *Sentencing: A Social Process* (Palgrave Macmillan, Cham, 2020) 119 at 135.

[119] Tata, above n 118, at 134.

[120] Tombs, above n 118, at 99.

[121] See Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st ed, Crown, 3 New York, 2016).

[122] Tombs, above n 118, at 99.

[123] AI Committee Members of Sabah and Sarawak, above n 6.

[124] AI Committee Members of Sabah and Sarawak, above n 6.

judiciously during sentencing hearing.[125] Such an approach is consistent with my suggestion of a new regulatory framework in the form of an Independent Monitor that will be explained in the final chapter.

Ultimately, when developing an algorithmic sentencing system, it is critical that the data collected is especially tailored to the needs of human judges. By cooperating with the end-user, such an approach could increase the usefulness of the algorithmic sentencing system. At the very least, with this approach, if the implementation of an algorithmic sentencing system in New Zealand courts were to fail, such failure is not due to the inability to meet the needs of the judiciary.

In the judicial context, an algorithmic sentencing system that would supplant human judges would not be desirable. A wholly mechanical sentence computation that converts a sentencing process to a mathematical and logistical exercise will jeopardise judicial autonomy and result in a loss of narrative, thus decontextualising and dehumanising sentencing which is considered as a form of art, requiring human insight, time, perseverance and experience.[126] It is in this sense that an algorithmic sentencing that generates non-binding recommendations would be more favourable. It will maintain not only judicial ownership of the sentencing process but uphold an individual's dignity. In Meg Lata Jones' own words, "to treat a human in a wholly computational manner reduces the individual's dignity and restoration of dignity can be provided by a human in the loop."[127]

C Conclusion

In this chapter, I have shown what an algorithmic sentencing system could look like in New Zealand. An algorithmic sentencing decision, envisioned by Chiao, is attractive but has its limitations. In particular, some cases involving more serious offences would require a larger number of variables that courts need to take into account. In such cases, there is more at stake

---

[125] AI Committee Members of Sabah and Sarawak, above n 6.

[126] Aas, above n 54, at 25.

[127] Meg Lata Jones "The right to a human in the loop: Political constructions of computer automation and personhood" (2017) 47(2) Social Studies of Science 216 at 232.

for offenders and the community alike. However, there is no reason technically and in principle that such a system cannot be designed to accurately compute the sentencing variables that are pertinent to the sentencing decision.

This chapter has identified three main concerns concerning an algorithmic sentencing system. These concerns are transparency, impartiality and judicial ambivalence. With any changes introduced to the status quo, challenges in terms of implementation are, reasonably, expected. What is required here, is an incremental approach requiring a thorough trial process with extensive consultations with relevant end-users and stakeholders, given the cardinal interests (political as well as social) that are involved in the legal domain of sentencing law.

It is important to clarify here that my proposal for an algorithmic sentencing system, drawing from Chiao's work, will only supplement and not supplant human judges in their roles of determining the appropriate levels of sentences. I contend that human sentencing is a judicial function, that should remain in the hands of humans. An algorithm sentencing system, based on a machine learning approach, will not be a satisfactory substitute for human decision-making but will act as an important tool in supporting judicial decision making in criminal sentencing by increasing efficiency and to some extent, consistency in the sentencing process.

CHAPTER III: SEEKING RECOURSE UNDER NEW ZEALAND'S CURRENT LEGISLATIVE FRAMEWORK

I will now explore the various recourses available under the existing law in New Zealand for individuals subject to algorithmic decisions. I will proceed based on a hypothetical situation where the Ministry of Justice, after extensive consultations with relevant stakeholders and rigorous trial sessions where the system is proven to produce highly accurate predictions, decided to introduce the algorithmic sentencing system mentioned in chapter II in New Zealand courts. This chapter will show that the current legislative framework is inadequate to provide legal remedies for those subject to algorithmic decisions, especially in terms of algorithmic transparency and the right against discrimination. At the end of the chapter, I will consider how New Zealand's Algorithm Charter could complement the new regulatory model proposed in chapter IV.

A Upholding Algorithmic Transparency via the Privacy Act, the OIA and the LGOIMA

The Privacy Act, OIA and LGOIMA confer informational rights relating to the transparency of algorithmic decisions, potentially helping those affected by the algorithmic predictions to challenge these decisions. For convenience's sake, when referring to the OIA, I am also referring to the LGOIMA. The following discussion will now briefly illustrate how the current legal avenues provide limited assistance in providing meaningful explanations to those individuals.

1 The Privacy Act

After a lengthy parliamentary process, the Privacy Act 2020 came into force on 1 December 2020, repealing and replacing the old Privacy Act 1993.[128] However, despite some significant changes to New Zealand's privacy law, the Act does not directly address the harms associated with algorithmic decision-making. Indeed, submissions have been made at the Select Committee stage for the need of explicit regulations of algorithmic decisions in the Privacy Bill 2018. These submissions were made on the basis to improve transparency and provide

[128] Privacy Commissioner "Privacy Act 2020" <https://privacy.org.nz/>.

individuals with greater autonomy over personal data used in these algorithmic decision-making processes.[129] However, the Selected Committee rejected such submissions[130] on the basis that current legislation adequately covers algorithmic decision-making and since "humans not computers are making almost all of the significant decisions affecting individuals in New Zealand", there is no pressing need for regulating the algorithmic system.[131] The following discussion will show how such assertions are flawed.

In terms of gaining access to information, Principle 6 of the Privacy Act (IPP6) allows individuals access to their personal information held by an agency. Here the Ministry of Justice falls under the definition of "agency" outlined in the Act.[132] Therefore, in principle, under IPP6, an individual can request any information about them used in a decision, including the input and output information relating to the algorithmic sentencing system. Consequently, an affected party may gain access to such information under IPP6 without having to prove harm.[133] Optimistically, access to such information could enable an individual to challenge the reasons behind an algorithmic decision. However, given the complex internal workings of machine learning algorithms, such a right may prove to be of limited practical assistance. The technical barriers in explaining how the sentencing system came to a decision meant that despite gaining access to the input and output data used for the algorithmic sentencing system, the affected party would not be able to understand how the algorithm prediction came to be, significantly limiting the right to correct, challenge and contest the information that has been used by the algorithmic system to influence the outcome of the sentencing decision.

2 The OIA and the LGOIMA

The OIA allows New Zealand citizens, permanent residents and anyone who is in New Zealand to request any official information held by government agencies - including the

---

[129] Privacy Commissioner John Edwards "Submission to the Justice and Electoral Select Committee on the Privacy Bill 2017" at 29, 30; Associate Professer Gehan Gunasekara "Submission to the Justice and Electoral Select Committee on the Privacy Bill 2017" at 3.

[130] Privacy Bill 2017 (34-2) (select committee report) at 39, 40.

[131] At 39, 40.

[132] Section 8.

[133] Section 31(2).

Ministry of Justice.[134] This right is subject to a few exceptions[135] and some parts of the information requested may be deleted.[136] It follows then that under the OIA, an affected party has a prima facie "right to reasons" for the sentencing decision made by the algorithmic sentencing system, provided that ss 27 and 32 do not apply. Both ss 27 and 32 cover good reasons to refuse access. The OIA sets out three requirements as to what a statement of reasons should include: These being the findings on material issues of fact; a reference to the information on which the findings are based; and the reasons for the decision or recommendation.[137]

The problem here is translating these statutory criteria into adequate reasons to justify the sentencing system's predictions. To allow an affected individual to effectively challenge the reasons behind an algorithmic decision, the reasons given must be "intelligible to the recipient"[138] and be of "sufficient precision to give him or her a clear understanding of why the decision was made".[139] Once again, given the technical complexities of machine learning algorithms, there is a risk that the aforementioned prima facie "right to reason" is rendered ineffective. If the reasons given do not comply with the aforementioned standards, it will prevent a fair assessment of the decision that has been made by an algorithmic sentencing system, thus consequently undermining judicial transparency.

3 Human in the loop: a satisfactory antidote to algorithmic decisions?

As previously mentioned, the Committee rejected the need for the regulation of algorithmic decision-making as humans are still involved in the decision-making process. I will now show

---

134 Ministry of Justice "Official Information Act Requests" < https://www.justice.govt.nz/>.

135 Sections 27 – 32.

136 Section 17.

137 Official Information Act 1982, s 23(1); and Local Government Official Information and Meetings Act 1987, s 22(1).

138 *Elliot v Southwark London Borough Council* [1976] 2 All ER 781, [1976] 1 WLR 499 (CA) at 508 per James LJ; adopted in *Re Palmer and Minister for the Capital Territory* (1978) 23 ALR 196 at 206–207 as cited in Graham Taylor and Paul Roth *Access to Information* (2nd ed, LexisNexis, 125 Wellington, 2011) at 194.

139 *Re Poyser and Mills' Arbitration* [1963] 1 ALL ER 612, [1964] 2 QB 467 at 478 per Megaw J; adopted in *Re Palmer and Minister for the Capital Territory,* above n 137 as cited in Taylor and Roth, above n 137, at 194.

how such a claim is problematic, arguing that even when humans are present in the decision-making process, additional oversight of these algorithmic-driven tools is still required.

Algorithmic systems can be used to varying extents together with human decision-making. In some scenarios, even if a human is present in the decision-making process, the predictions generated by an algorithmic system can still have a heavy influence on the human decision-maker. In our hypothetical scenario, given the high accuracy showed by the algorithmic sentencing system based on previous trial sessions, a judge will likely put heavy reliance on the predictions made by the algorithmic sentencing system, even against their better judgment. This phenomenon is usually referred to as automation bias which occurs when humans tend to unthinkingly defer to an automated decision.[140] Indeed, several studies have shown that there is an over reliance on automated decisions by human decision-makers.[141] For instance, algorithmic decision-making is heavily relied upon in the aviation section, where 40% of pilots over-relied on an automated solution to a flight route configuration, using none of their own reasoning and accepting suboptimal flight plans.[142] Furthermore, if the algorithmic system is discriminatory in any way, human decisions relying on the algorithmic system might replicate the bias. Since human decision-makers do not have access to the list of rejected data and do not have enough information to detect any important information that the algorithmic systems might have missed, the presence of bias or inaccuracy will be difficult to spot and mitigate, possibly leading to unfair results.

B Challenging Algorithmic Bias via the HRA

This section will now consider the extent to which the HRA could respond adequately to algorithmic bias. It will show that an affected party will have the best chance of establishing

---

[140] Ella Brownlie "Encoding Inequality: The Case for Greater Regulation of Artificial Intelligence and Automated Decision-Making in New Zealand" (2020) 51(1) VUWLR 1 at 16.

[141] Eugenio Alberdi, Lorenzo Strigini, Andrey A. Povyakalo, Peter Ayton "Why are People's Decisions Sometimes Worse with Computer Support?" in B Buth, G Rabe, T Seyfarth (eds.) *Computer Safety, Reliability and Security* included in *Lecture Notes in Computer Science*, vol 5775 (Springer, Berlin, 2009) at 1.

[142] ML Cummings (MIT) "Automation Bias in Intelligent Time Critical Decision Support Systems" (paper presented to American Institute of Aeronautics and Astronautics 1st Intelligent Systems Technical Conference 20-22 September 2004) <https://arc.aiaa.org/doi/abs/10.2514/6.2004-6313> at 4.

prima facie discrimination under an algorithmic directed decision, i.e., where the decision is fully automated. However, an affected party will have a lesser chance to succeed if subject to an algorithmically informed decision, i.e., where the algorithmic output is just one of the factors considered by human decision-makers. Furthermore, given the individualistic framework of the HRA, it is unlikely that the Act can respond adequately to indirect discrimination resulting from the use of an algorithmic system.

Section 19 of the New Zealand Bill of Rights Act, 1990 gives everyone in New Zealand the right to freedom from discrimination. The HRA, which came into force on 1 February 1994, deals with discrimination by providing remedies to those discriminated against, based on a prohibited ground of discrimination ("prohibited ground"). Section 21 of the HRA sets out a list of the prohibited grounds for discrimination. An affected party can bring a discrimination claim based on a public sector agency's actions under Part 1A of the HRA. To prove that an algorithmic decision has caused unjustified discrimination, the affected party will need to show that:[143] There is differential treatment or effects as between persons or groups in analogous or comparable situations on the basis of a prohibited ground; the differential treatment has a discriminatory impact, that is, when "viewed in context, it imposes a material disadvantage on the person or group differentiated against"; in accordance with section 5 of New Zealand Bill of Rights Act, and the differential treatment is not a limitation on the right to be free from discrimination found in s 19 of New Zealand Bill of Rights Act that "can be demonstrably justified in a free and democratic society".

The challenges here lie in proving a prima facie case for those subject to algorithmically informed decisions and proving indirect discrimination. Pertaining to the former, for those affected by algorithmically informed decisions, the challenge here is to prove that the discriminatory output is a "material ingredient" or "operative factor" in the differential treatment. Applying this to an algorithmic sentencing system, if a human judge unquestioningly implements an algorithmic prediction imposing different treatment based on a prohibited ground, differential treatment will be easy to prove. However, given the advisory nature of the algorithmic sentencing system, if the prediction is only one of the many factors

[143] *Ministry of Health v Atkinson* [2012] NZCA 184, [2012] 3 NZLR 456 at [55], [109] and [143]; followed in *Ngaronoa v Attorney-General*; *Taylor v Attorney-General* [2017] NZCA 351, [2017] 3 NZLR 643.

relied on by the human judge, a problem arises of whether the prediction was a material ingredient in deciding the appropriate level of sentencing. The harm resulting from such a situation can also be seen in *Loomis*. In that case, the court considered that the output generated by the predictive tool was not "determinative" as it was only of the many criteria taken into account by the judge.[144] The decision seems problematic because the court here fails to take into account the possibility that the output might have a material influence on the sentencing outcome and the decision is also inconsistent with the vast literature on automation bias, which suggests judges could unthinkingly defer to algorithms output. Nevertheless, the case does illustrate that in the New Zealand legal landscape, the extent to which the human decision-maker retains discretion and control is important in determining whether the algorithmic output has materially resulted in the differential treatment, which proves to be one of the key challenges in bringing a claim under the HRA for those subject to algorithmic informed decisions.

Even prior to proving materiality and causation, an affected party, whether under an algorithmic directed decision or algorithmically informed decision will have a challenge to bring a claim of indirect discrimination. It is important to note that for algorithmic decision-making, most claims will typically be an indirect discrimination claim. An indirect discrimination claim arises not because the output generated by the algorithm system is on its face discriminatory due to inclusion of the protected characteristic such as race itself but rather due to unintentional reasons relating to the algorithm's model or data sources, the algorithm provides output which discriminates against a particular group on a prohibited ground.[145] For instance, as mentioned in chapter II, certain variables like a person's postcode used as input in an algorithmic system may serve as a proxy for race and indirectly discriminate. Establishing that there has been discrimination as to a person's race proves to be difficult here as the complainant cannot do so by reference to his own information but will need to have access to all decisions made to compare whether those of his ethnicity are more likely to be discriminated (i.e., more likely to get a harsher sentence of imprisonment) compared to those from different ethnicities. The lack of precedent to bolster an indirect discrimination case also

[144] *Loomis,* above n 101, at [102] - [110].

[145] Smith, above n 4 at 68.

proves to be a challenge. While the HRA clearly provides an avenue for bringing an indirect discrimination claim, none has been successful.[146] *Ngaronoa v Attorney-General* may provide some assistance but it shows there is a high bar for an indirect discrimination case to succeed.[147]

C New Zealand's Algorithm Charter

Recognising the risks associated with algorithmic decisions, the New Zealand government recently published the Algorithm Charter for Aotearoa New Zealand to increase public trust in the safe and effective use of data across government agencies.[148] Claimed as an international first, the Charter outlines a set of standards to guide the use of algorithms by public agencies.[149]

The Charter has so far been signed by 26 government agencies, including the Ministry of Justice.[150] The commitments under the Charter include maintaining transparency by clearly explaining how decisions are informed by algorithms and this may include plain English documentation of the algorithm, embedding a Te Ao Māori perspective in the development and use of algorithms consistent with the principles of the Treaty of Waitangi, consulting with people, communities and groups who have an interest in algorithms and likely to be impacted by their use, providing a channel for challenging or appealing of decisions informed by algorithms and clearly explaining the role of humans in decisions informed by algorithms.[151] In signing the Charter, Charter signatories will assess their algorithm decisions, using a risk matrix approach.[152] This risk matrix approach supports their evaluation, by quantifying the likelihood of an unintended adverse outcome against its relative level of impact to derive an

---

[146] Successful cases are usually related to issues concerning access to funding for parents supporting their disabled children, and resourcing cuts for services aimed at intellectually disabled persons over 65. See *Ministry of Health v Atkinson*, above n 142; and *Attorney-General v IDEA Services Ltd (In Statutory Management)* [2012] NZHC 3229, [2013] 2 NZLR 512.

[147] *Ngaranoa v Attorney General*, above n 142.

[148] New Zealand Government "New Algorithm Charter a world first" (28 July 2020) Beehive.govt.nz <https://www.beehive.govt.nz/>.

[149] New Zealand Government, above n 148.

[150] Stats NZ, above n 45.

[151] Stats NZ, above n 45.

[152] Stats NZ, above n 45.

overall level of risk.[153] The intention is to focus on those uses of algorithms that have a high or critical risk of unintended harms for New Zealanders.[154]

Although the Charter established a more principled approach to algorithmic decisions and could serve as a blueprint for other governments,[155] at the moment, it is still very high level and presently contains no mechanism for compliance checking. Therefore, I submit that there still needs to be a regulatory body that should have "an all-of-government remit".[156] In this respect, the Independent Monitor could be a good way of filling that gap. Government agencies could have to report every year to the Monitor, explaining how they have used algorithms and how they have made sure that the use of algorithms complies with the law and the key commitments outlined in the Charter.

D Conclusion

This chapter has shown that the current legislative framework is inadequate to respond to the algorithmic harms mentioned in chapter II. The Privacy Act and the OIA, while allowing individuals to gain access to information relating to the algorithmic system, prove to be of limited use given the technical barriers often associated with machine learning algorithms. Without having a meaningful understanding of the information provided, an affected party will not be able to properly challenge the reasons behind the algorithmic decisions. The HRA proves to be of limited assistance as well, especially when it comes to algorithmically informed decisions. This is illustrated in the case of *Loomis*. While the Algorithm Charter sets a strong foundation for guiding government agencies on how to use algorithmic systems in a manner that warrants public trust, the standards outlined do not include an enforcement mechanism, and this is where a new regulatory framework in the form of an Independent Monitor would be helpful.

---

[153] Stats NZ, above n 45.

[154] Stats NZ, above n 45.

[155] Charlotte Graham-McLay in Wellington "New Zealand claims world first in setting standards for government use of algorithms" *The Guardian* (online ed, London, 27 July 2020) quoting Professor Colin Gavaghan.

[156] Graham-McLay, above n 155.

CHAPTER IV: A NEW REGULATORY FRAMEWORK

The following discussion will now look to the GDPR for inspiration. While this chapter concludes that the GDPR is inadequate to address risks associated with algorithmic decisions, the European Commission's recent proposal for AI Regulation looks more promising in addressing the algorithmic harms mentioned in chapter II as it contains a more comprehensive regulatory compliance. I will then explore the available options to strengthen the current legislative framework, which would complement the new regulatory model I propose at the end of this chapter. Through a new regulatory model, I argue that a broader approach, is required to mitigate the harms associated with algorithmic sentencing decisions rather than relying on an individualistic framework of the current law in New Zealand.

A The GDPR as an Exemplar Legislation?

The GDPR is a regulation in European Union Law on data protection and privacy.[157] At the heart of GDPR is personal data, which allows a living person to be directly or indirectly identified from the available data.[158] The regime contains 99 articles, but arts 13, 15 and 22 are often discussed by academics regarding their usefulness in giving meaningful remedy to those subject to algorithmic decision making.

On plain reading, an affected party is afforded a right to an explanation under arts 13 and 15. However, whether such a right is meaningful in terms of allowing the affected party to understand the reasons behind an algorithmic decision remains contentious. Under art 13, an organisation using an algorithmic system will be required to notify affected individuals of such use and give them "meaningful information about the logic involved as well as the significance and envisaged consequences of processing for the data subject". Meanwhile, art 15 states that individuals should have a right to access the same information. In order to provide meaningful recourse to an affected individual, what is required here is an ex-post explanation about a specific decision that will allow an individual to challenge such a decision. While some scholars

---

[157] Liz Blythe, Joe Edwards and Rachel O'Brien "GDPR – A summary" (4 May 2018) Russell McVeagh <https://www.russellmcveagh.com/>.

[158] General Data Protection Regulation 2016 art 4(1), art 4(4).

are more generous in their interpretations of the articles to include an ex-post explanation,[159] current interpretations illustrate that arts 13 and 15 do not exist to provide such explanations, rather the explanation required by arts 13 and 15 is merely an ex ante explanation.[160] As a result, the right to explanation provided by arts 13 and 15 simply offers basic information about how the algorithm decision will be made and the expected consequences, rather than a detailed explanation of the rationale of the decision that would allow an individual to contest it.

Article 22 of the GDPR covers automated individual decision-making, including profiling. It states that "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling which produces legal effects concerning him or her or similarly significantly affects him or her". However, three exceptions apply to this general prohibition. These exceptions include: Where the decision is necessary for entering into or performing a contract, where the decision is authorised by Member state law and where the decision is based on the individual's explicit consent.[161] Previously misinterpreted as a right to explanation of automated decisions, art 22 is now affirmed by legal experts to be a prohibition clause on the use of automated decision making in some contexts.[162]

The issue arises here with regard to the interpretation of the word "solely". For art 22 to apply, the decision made must be "based solely on automated processing". This was defined as a decision that would be made "without human involvement"[163] and does not cover superficial human involvement. In this respect, the general prohibition does not apply where there is "meaningful oversight" by humans with sufficient authority and competency to change the

---

[159] Bryce Goodman and Seth Flaxman "European Union Regulations on Algorithmic Decision-making and a Right to Explanation" (2017) 38 AI Magazine 50.

[160] Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' is probably not the remedy you are looking for?" (2017) 16 Duke Law and Technology Review 1 at 52.

[161] Article 9(1).

[162] Sandra Wachter, Brent Mittelstadt and Luciano Floridi "Why a Right to Explanation of Automated Decision Making Does Not Exist in the General Data Protection Regulation" 2017 7(2) International Data Privacy Law 76 at 77 – 78.

[163] Article 29 Data Protection Working Party "Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679" at 20.

results made by the algorithm system.[164] This is consistent with the assumption by the Select Committee on the Privacy Bill in New Zealand that as long as there are humans involved when using the algorithmic system, the harms associated with algorithmic decision-making can be prevented.[165] Once again, the case of *Loomis* illustrates how such an assumption is problematic. The judge's decision in *Loomis* will be unlikely to fulfil the definition of "solely automated" under art 22 as the judge was not strictly bound to the score but rather is free to take other considerations into account. As discussed earlier, keeping humans in the loop will not help to ameliorate the harms associated with algorithmic decision-making as it fails to consider the presence of automatic complacency and the lack of human oversight over the algorithmic tool themselves.

B A New Dawn of International AI Regulation?

On April 21, 2021, the European Commission published its much-anticipated proposal for a regulatory framework to monitor AI.[166] Considered the first-ever legal framework on AI in the European Union, the AI Regulation establishes rules regarding the development, placement on the market and the use of AI systems, with hefty penalties for non-compliance. Although still at its draft stage, if adopted, the AI Regulation would have serious consequences for agencies (public and private) who develop, sell or use AI systems.

AI systems are defined in Annex I of the AI Regulation. Those include, for example, machine learning (ML), logic or knowledge based and statistical approaches and can, for a given set of human-defined objectives, generate output such as content, predictions, recommendations, or decisions influencing the environments they interact with.[167] Under the AI Regulation, the Commission proposes a "risk-based" approach by classifying AI systems according to the risk

---

[164] At 21.

[165] Privacy Bill 2017 Select Committee Report, above n 130, at 39,40.

[166] European Commission *Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* (2021/0106 (COD), 21 April 2020).

[167] Article 3(1).

they pose to human beings.[168] At its heart, the AI Regulation focuses on identifying and monitoring high-risk AI systems. In this section, the discussion will focus on the obligations imposed on high-risk AI systems under the AI Regulation.

High-risk AI systems include AI technology that is used in the administration of justice and democratic processes.[169] The algorithmic sentencing system envisioned in chapter II falls under that category. Under the AI Regulation, the algorithmic sentencing system will be subject to several obligations before it is introduced in court and throughout its life cycle. In terms of pre-deployment obligations, before placing a high-risk AI system on the European Union market or otherwise putting it into service, the AI Regulation requires the AI system to undergo a conformity assessment.[170] Such an assessment is necessary to demonstrate that the AI system complies with the mandatory requirements for trustworthy AI related to data quality, documentation and traceability, transparency, human oversight, accuracy, and robustness.[171] In addition, the assessment will have to be repeated in the event that the system itself or its purpose has been significantly changed. For some AI systems, an independent notified body will also participate in this process. Post-market monitoring includes quality and risk management systems, audits and reports on serious incidents or breaches of fundamental rights obligations.[172] Violations of the AI Regulation attract potentially significant administrative fines, ranging from 2-6% of worldwide annual turnover, depending on the violation.[173]

The AI Regulation, I submit, is a significant legislative attempt that could serve as a model for regulating AI systems in New Zealand, especially the algorithmic sentencing system I envision in chapter II. It provides a sound-risk based structure and creates a new, holistic framework

---

[168] European Commission "Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence" (21 April 2021) European Commission <https://ec.europa.eu/info/index_en>.

[169] Annex III (8).

[170] Article 19.

[171] Articles 10 – 15.

[172] European Commission, above n 166.

[173] Article 71.

of regulators and testing, monitoring and compliance processes. Moreover, its cradle-to-grave approach will ensure that the harms associated with an algorithmic sentencing system are mitigated before its deployment in courts and throughout its lifecycle.

C Improving the Current Legislative Framework in New Zealand

I will now look at the available options that could improve the current law in New Zealand. These options, if implemented, will serve to supplement the new regulatory model proposed at the end of this chapter.

1 Explainable AI

The government can either choose to implement machine learning algorithms that are explainable by design or instead delay the deployment of such algorithms until they can generate an explanation that sufficiently explains the basis of their outcome. This is where "Explainable AI" or "XAI" comes into play. Described as "any machine learning technology that can accurately explain a prediction at the individual level", Explainable AI, while still under development, will help affected individuals to understand a decision made by an algorithm system.[174] If questions such as "How does the algorithm system work?", "What mistakes can the system make?" and "Why did the system just do that?" can be answered meaningfully, the goal of explanation is satisfied.[175] This solution looks even more promising with the express commitment of large technologies such as Google, IBM and Microsoft to develop an interpretable machine learning system.[176] Facebook, for instance, has implemented a partial use of Explainable AI, with the introduction of a "Why am I seeing this?" feature that informs its users why they see a particular advertisement or post on their feed.[177] However, a trade-off will exist here where producing explainable systems would typically require reducing

---

[174] "Explainable AI" simMachines <https://simmachines.com>.

[175] Expert.ai Team "Explainable Artificial Intelligence Explained" (5 November 2020) expert.ai <https://www.expert.ai/?>.

[176] Charles Towers-Clark "Can We Make Artificial Intelligence Accountable?" *Forbes* (online ed, United States of America, 19 September 2018).

[177] Alex Hern "Why am I seeing this? New Facebook tool to demystify Newsfeed" *The Guardian* (online ed, United Kingdom, 1 April 2019).

the system's complexity and thus affecting its accuracy.[178] As mentioned in chapter II, this type of policy decision-making is one of the many decisions that will be confronted under the new regulatory model.

2 Guidelines for standard of explainability

In assessing the standard of explainability required for algorithmic decision-making, it is useful to compare it to the standard of explainability required for human decision making. Given that human decision-making is far from perfect, what we are seeking here are practical explanations, not higher than expected of human decision makers, that would allow an affected party or a review tribunal to make a full and fair assessment of the decision that has been made with the support of an algorithmic sentencing system. Perhaps Canada's Directive on Automated Decision-Making serves as a useful starting point. Under the Canada's Directive, for very low impact decisions, plain-language "frequently asked questions" about an algorithm's processes, supported by a general non-technical description of the reasoning for the decision at hand, might be considered adequate.[179] However, for more high-stakes decisions, reasons could require: A more detailed technical description of how the model works; the nature of the training data used; how the model has been validated and any relevant audits or reviews that have been undertaken; and information about how the model contributed to the decision at hand. Given the nature of criminal sentencing, the latter category would be preferred.

D A New Regulatory Model for Algorithmic Systems

Previous discussions demonstrate how the current legislative framework, which adopts an individualistic framework focusing on right-based claims, is inadequate to address potential risks relating to algorithm decision-making. Thus, I argue the need for a new regulatory model in the form of an Independent Monitor, which is tasked with, among other things, ensuring

---

[178] John Zerilli, Alistair Knott, James Maclaurin, Colin Gavaghan "Transparency in Algorithmic and Human Decision Making: Is there a double standard?" (5 September 2018) Philosophy and Technology <https://doi.org/10.1007/s13347-018-0330-6> at 4.

[179] Canadian Government Directive on Automated Decision-Making (April 2019, Appendix C – Impact Level Requirements.

compliance with algorithmic impact assessments (AIAs), as such a model would better address the algorithmic harms mentioned in chapter II. This model proposed by several experts in this area does not only help to ameliorate the concerns related to algorithmic sentencing systems but also uses of the algorithm system in other contexts. In addition, this top-down approach is more likely to address some of the more collective and diffuse harms related to an algorithmic sentencing system that exists beyond an individual level.

Under the Independent Monitor approach, this regulatory body would have different funding and be statutorily independent of the existing government, removing the influence of political factors that could serve as obstacles to the development of an algorithmic system. Furthermore, this regulatory body would encourage best practice and use of algorithmic systems by mandating compliance for AIAs. AIAs work much like privacy impact assessments, which will be used to evaluate any proposed algorithmic use cases and detect the potential risks that arise from the proposed use, allowing to put practical mitigations in place.[180] With the power afforded to the Independent Monitor, they can request information from public sector agencies that want to employ the algorithmic system and carry out auditing if necessary. To promote public disclosure, the Independent Monitor can report annually to the Parliament on the use of algorithm systems in any government agency and disclose any risks or problems related to the algorithmic sentencing system, enhancing transparency and accountability.[181] The Independent Monitor would also be tasked with carrying an external review to detect any adverse impacts with regard to the use of the algorithmic system.[182] As mentioned in chapter III, the Algorithm Charter could complement this new regulatory model as it outlines several key commitments associated with the use of algorithmic systems for government agencies. In addition, in many aspects, the European Commission's proposed AI Regulation may also serve as a useful guidance especially the requirements where AI systems have to undergo a mandatory pre-deployment accuracy report before being used in their relevant sectors. In addition, as suggested by Gavaghan et al, consultations with populations likely to be affected by algorithmic decisions and with those likely to be under-represented in construction and

[180] AI Now Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability (April 2018) at 9.

[181] Smith, above n 4, at 107.

[182] At 107.

training of algorithmic systems should also be conducted as these individuals are likely to have valuable input that may not be thought of even by the most well-intentioned observers.[183]

A hard-edged regulatory model could be an alternative option. However, with stronger enforcement powers, such an approach might increase resourcing costs and likely discourage agencies from self-reporting defects in the algorithmic sentencing system. Furthermore, such an approach is arguably more appropriate in regulating the conduct of private sectors that are not subject to political accountability. A self-regulation approach might also be attractive as it will attract less cost compared to an Independent Monitor and provides more flexibility. However, it is likely to promote inconsistencies in processes and standards across government generally.[184]

E Conclusion

The legal avenues for those affected by a decision made by an algorithm may be available in a range of guises under the current New Zealand law, including the Privacy Act, OIA and the HRA. However, as this chapter has shown, the remedies available under each legislation provide limited practical assistance especially considering the complex inner workings of machine learning algorithms. However, I submit that New Zealand's Algorithm Charter could complement the new regulatory model as it outlines a set of standards to guide government agencies in using and developing algorithm systems. While the GDPR also proves inadequate in addressing harms relating to limitations mentioned in chapter II, the recent AI Regulation proposed by the European Commission may serve as an inspiration for regulating the algorithmic decisions in New Zealand with its cradle-to-grave approach. While the current legislative framework in New Zealand can be strengthened with other available solutions, a new regulatory body in the form of an Independent Monitor is necessary to promote the ethical and effective use of algorithmic system in New Zealand government agencies, especially in the Ministry of Justice where sentencing decisions have greater and significant effects on an individual's fundamental rights and liberty.

---

[183] Gavaghan, Knott, Maclaurin, Zerilli and Liddicoat, above n 4, at 77.

[184] Tutt, above n 4, at 114.

CONCLUSION

This dissertation aims to make a modest case for introducing an algorithmic sentencing system in New Zealand. It contends that an algorithmic sentencing system will serve as a helpful tool to assist judges in determining an appropriate sentence by generating non-binding sentence predictions and systemising past sentencing decisions to inform the judge of how their previous colleagues have reacted. It fully acknowledges the complexities and the discretionary nature of sentencing, an area of law that has traditionally been in the domain of human judges. However, as this dissertation has shown, there are serious flaws in the current sentencing system. While an algorithmic sentencing system may not be the panacea for all the problems in the sentencing process, especially broader systemic issues that occur beyond the criminal justice system, it argues that such a system would improve sentencing primarily in terms of efficiency and to some extent, consistency.

As reminded throughout this dissertation, when assessing the effectiveness of an algorithmic sentencing system to improve sentencing law, the issue is comparative – we must measure the system against the currently flawed context it will operate in instead of a perfect or ideal system. Furthermore, when introducing a change to the status quo, we can reasonably expect challenges throughout all stages of its implementation. In the case of an algorithmic sentencing system, what is required is a cradle-to-grave approach whereby the technology must be evaluated even before it is put into service and throughout its lifecycle. This approach ensures that the fundamental rights and cardinal interests of those subject to algorithmic decisions are protected. Of course, there exists no perfect solution to a problem, but it is hoped that the new regulatory framework in the form of an Independent Monitor will adequately address the concerns associated with algorithmic sentencing decisions such as transparency, impartiality and judicial ambivalence, with New Zealand's Algorithm Charter and the European Commission's proposed AI Regulation setting a solid foundation.

BIBLIOGRAPHY

A Cases

1 New Zealand

Attorney-General v IDEA Services Ltd (In Statutory Management) [2012] NZHC 3229, [2013] 2 NZLR 512.

Ministry of Health v Atkinson [2012] NZCA 184, [2012] 3 NZLR 456.

Ngaronoa v Attorney-General; *Taylor v Attorney-General* [2017] NZCA 351, [2017] 3 NZLR 643.

R v Morris [1991] 3 NZLR 641 (CA).

R v Taueki [2005] 3 NZLR 372 (CA).

2 England and Wales

Elliot v Southwark London Borough Council [1976] 2 All ER 781, [1976] 1 WLR 499 (CA).

Re Poyser and Mills' Arbitration [1963] 1 ALL ER 612, [1964] 2 QB 467.

Re Palmer and Minister for the Capital Territory (1978) 23 ALR 196.

3 United States

State v Loomis 881 NW 2d 749 (Wis 2016).

B Legislation and Bills

1 New Zealand

Human Rights Act 1993.

Local Government Official Information and Meetings Act 1987.

New Zealand Bill of Rights Act 1993.

Official Information Act 1982.

Privacy Act 2020.

Sentencing Act 2002.

Privacy Bill (34-2).

Privacy Bill 2017 (34-2) (select committee report).

2 European Union

General Data Protection Regulation 2016.

C Books

Katja Franko Aas *Sentencing in the Age of Information: from Faust to Macintosh* (The GlassHouse Press, London, 2005).

Bernard E. Harcourt *Against Prediction: Profiling, Policing, and Punishing in An Actuarial Age* (University of Chicago Press, Chicago, 2006).

David J. *Gunkel Gaming the System: Deconstructing Video Games, Game Studies, and Virtual Worlds* (Indiana University Press, Bloomington, 2018).

Geoff Hall *Sentencing Law and Practice* (LexisNexis, Wellington, 2004).

Geoff Hall *Sentencing: 2007 Reforms in Context* (Lexis Nexis, Wellington, 2007).

Daniel Kahneman *Thinking, Fast and Slow* (1st ed, Daniel Kahneman, New York, 2017).

Taylor R Moore *Trade Secrets and Algorithms as Barriers to Social Justice* (Centre for Democracy & Technology, Washington, 2017).

Cathy O'Neil *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st ed, Crown, 3 New York, 2016).

Frank Pasquele *The Black Box Society: The Secrets Algorithms That Control Money and Information* (Harvard University Press, Cambridge, 2015).

John Rawls *A Theory of Justice* (Oxford University Press, Oxford,1999).

Graham Taylor and Paul Roth *Access to Information* (2nd ed, LexisNexis, 125 Wellington, 2011).

Jacob Turner *Robot Rules: Regulating Artificial Intelligence* (Springer, Switzerland, 2019).

D Chapters in Books

Eugenio Alberdi, Lorenzo Strigini, Andrey A. Povyakalo, Peter Ayton "Why are People's Decisions Sometimes Worse with Computer Support?" in B. Buth, G. Rabe, T. Seyfarth (eds.) *Computer Safety,Reliability and Security* included in *Lecture Notes in Computer Science*, vol 5775 (Springer, Berlin, 2009).

Gabrielle Appleby and Heather Roberts "The Chief Justice: Under Relational and Institutional Pressure" in Gabrielle Appleby and Andrew Lynch *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, Cambridge, 2021).

Richard Devlin and Adam Dodek "Regulating Judges: Challenges, Controversies and Choices" in Richard Devlin and Adam Dodek *Regulating Judges: Beyond Independence and Accountability* (Edward Elgar Publishing, Cheltenham, 2016).

Julian V. Roberts and Mojca M. Plesnicar "Sentencing, Legitimacy and Public Opinion" in Gorazd Mesko and Justice Tankebe *Trust and Legitimacy in Criminal Justice: European Perspectives* (Springer, Cham, 2015).

Jacqueline Tombs "Telling sentencing stories" in Pat Carlen *Imaginary Penalties* (1st ed, Willan Publishing, Cullompton, 2008).

Cyrus Tata "The Application of Judicial Intelligence and "Rules" to Systems Supporting Discretionary Judicial Decision-Making" in G. Sartor and L. Karl Branting *Introduction: Judicial Applications of Artificial Intelligence* (Kluwer Academic Publishers, Dordrecht, 1998).

Cyrus Tata "The Rise of Technology and the Demise of the Sentencing Professions?" in Dave Cowan *Sentencing: A Social Process* (Palgrave Macmillan, Cham, 2020).

Monika Zalnieriute and Felicity Bell "Technology and Judicial Role" in Gabrielle Appleby and Andrew Lynch *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia* (Cambridge University Press, Cambridge, 2021).

E Journal Articles

Katja Franko Aas "From narrative to database: Technological change and penal culture" (2004) 6(4) Punishment & Society 379.

Nikolaos Aletras, Tsarapatsanis Dimitrios, Preoţiuc-Pietro Daniel, and Lampos Vasileios "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective" (2016) 2 PeerJournal of Computer Science 93.

Felicity Bell "Family Law, Access to Justice, and Automation" (2019) 19 Macquarie L.J. 103.

Kiel Brennan-Marquez and Vincent Chiao "Algorithmic Decision-Making When Humans Disagree on Ends" (2021) 24(3) New Criminal Law Review 275.

Ella Brownlie "Encoding Inequality: The Case for Greater Regulation of Artificial Intelligence and Automated Decision-Making in New Zealand" (2020) 51(1) VUWLR 1.

Oren Ben-Dor "The institutionalisation of public opinion: Bentham's proposed constitutional role for jury and judges" (2007) 35 27 Legal Stud 216.

The Honourable TF Bathurst AC "Who judges the judges, and how should they be judged?' (2019) 14 TJR 19.

Vincent Chiao "Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice" (2019) 15(2) Int JLC 126.

Vincent Chiao "Predicting Proportionality: The Case for Algorithmic Sentencing" (2018) 37(3) Crim Just Ethics 238.

Ryan Calo "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51 UC Davis L Rev 399.

Anupam Datta, Matt Fredrikson, Gihyuk Ko, Piotr Mardziel, Shayak Sen "Proxy Discrimination in Data-Driven Systems: Theory and Experiments with Machine Learnt Programs" (2017) Cornell University.

Michael E. Donohue "A Replacement for Justitia's Scales? Machine Learning's Role in Sentencing" (2019) 32(2) Harv JL & Tech 658.

Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' is probably not the remedy you are looking for?" (2017) 16 Duke Law and Technology Review 1.

Lilian Edwards and Michael Veale "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017-2018) 16 Duke L & Tech Rev 18.

Bryce Goodman and Seth Flaxman "European Union Regulations on Algorithmic Decision making and a Right to Explanation" (2017) 38 AI Magazine 50.

Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi "A survey of Methods for Explaining Black Box Models" (2018) 51(5) ACM Comput Surv 1.

Wayne Goodall and Russil Durrant "Regional Variation in Sentencing: The incarceration of aggravated drink drivers in the New Zealand District Courts" (2013) 46 ANZJ Crim 422.

HLA Hart "Positivism and the Separation of Laws and Morals" (1958) 71(4) Harv L Rev 593.

Kelly Hannah-Moffat "Actuarial Sentencing: An 'Unsettled' Proposition" (2012) 30(2) JQ 270.

Bettina von Helversen and Jorg Reiskamp "Predicting sentencing for low-level crimes: Comparing models of human judgment" (2009) 15 Journal of Experimental Psychology: Applied 375-395.

Meg Lata Jones "The right to a human in the loop: Political constructions of computer automation and personhood" (2017) 47(2) Social Studies of Science 216.

Samantha Jeffries, Garth Fletcher and Greg Newbold "Pathways to Sex-Based Differentiation in Criminal Court Sentencing" (2006) 41 Criminology 329.

Andreas Kaplan and Michael Haenlein "Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence" (2020) 63(1) Business Horizons 37.

Sarah Krasnostein and Arie Freiberg "Pursuing Consistency in an Individualist Sentencing Framework: If You Know Where You're Going, How Do You Know When You've Got There?" (2013) 76 LCP 265.

Bruno Lepri, Nuria Oliver, Emmanuel Letouze, Alex Pentland and Patrick Vinck "Fair, Transparent and Accountable Algorithmic Decision-Making Process" (2017) 31(4) Philos Technol 612.

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) I & CTL 223.

Jesper Ryberg "Sentencing Disparity and Artificial Intelligence" The Journal of Value Inquiry 1.

Andrea Roth "Trial by Machine" (2016) 104(5) Geo LJ 1245.

Carla L Reyes and Jeff Ward "Digging into Algorithms: Legal Ethics and Legal Access" (2020) 21 Nev LJ 325.

Sonja B. Starr "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination" 66 *Stan L Rev* 803.

Ric Simmons "Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System" (2018) 52(2) UC Davis L Rev1067.

Matthew U. Scherer "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies" (2016) 29(2) Harv JL & Tech 353.

Richard E. Susskind "Expert Systems in Law: A Jurisprudential Approach to Artificial Intelligence and Legal Reasoning" (1986) 49(2) MLR 168.

Harry Surden "Artificial Intelligence and Law: An Overview" (2019) 35 Ga St UL Rev 1305.

Harry Surden "The Variable Determinacy Thesis" (2011) 12 Colum Sci & Tech L Rev 1.

JJ Spigelman, "The quality dimension of judicial administration" (1999) 4(3) TJR 179.

Andrew Tutt "An FDA for Algorithms" (2016) 69 Admin L Rev 83.

Sandra Wachter, Brent Mittelstadt and Luciano Floridi "Why a Right to Explanation of Automated Decision Making Does Not Exist in the General Data Protection Regulation" 2017 7(2) International Data Privacy Law 76.

Monika Zalnieriute, Lyria Bennett Moses and George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82(3) MLR 425.

F Conference and Seminar Papers

M.L. Cummings (MIT) "Automation Bias in Intelligent Time Critical Decision Support Systems" (paper presented to American Institute of Aeronautics and Astronautics 1st Intelligent Systems Technical Conference 20-22 September 2004) <https://arc.aiaa.org/doi/abs/10.2514/6.2004-6313>.

G Unpublished Papers

Briony Blackmore "Developing Transparency Requirements for the Operation of Criminal Justice Algorithms in New Zealand" (PhD Thesis, University of Otago, 2019).

Wayne Goodall "Sentencing Consistency in the New Zealand District Courts" (PhD Thesis, Victoria University of Wellington, 2014).

David Smith "The Citizen and the Automated State: Exploring the Implications of Algorithmic Decision-making in the New Zealand Public Sector" (Master's Thesis, Victoria University of Wellington, 2020).

Frej Klem Thomsen "Iudicium ex Machinae – The Ethical Challenges of Automated Decision-Making in Criminal Sentencing" (October 2020 draft, Danish Institute for Human Rights, Copenhagen, 2020).

Jessie Malcolm "Exploring the Enigma: Enhancing Digital Rights in the Age of Algorithms" (LLB (Hons) Dissertation, University of Otago, 2018).

H Reports and Guidance

AI Now Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability (April 2018).

Article 29 Data Protection Working Party "Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679".

Colin Gavaghan, Alistair Knott, James Maclaurin, John Zerilli and Joy Liddicoat *Government Use of Artificial Intelligence in New Zealand: Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project* (Wellington, 2019).

Canadian Government Directive on Automated Decision-Making (April 2019, Appendix C – Impact Level Requirements.

European Commission *Proposal for a Regulation of the European Parliament and of the Council: Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.* (2021/0106 (COD), 21 April 2020).

I Newspaper and Magazine Articles

Jordan Bond "Maori Imprisoned at Twice Rate of Europeans for the Same Crime" *NZ Herald* (online ed, New Zealand, 14 September 2016).

Chris Baranjuk "Durham Police AI to help with custody decisions" *BBC* (online ed, London,10 May 2017).

Emile Donovan "Delayed Justice – Our Courts Under Pressure" *Newsroom* (online ed, New Zealand, 28 July 2020).

Charlotte Graham-McLay in Wellington "New Zealand claims world first in setting standards for government use of algorithms" *The Guardian* (online ed, London, 27 July 2020).

Jendy Harper "Why Does NZ Imprison So Many Maori" *Newsroom* (online ed, New Zealand, 29 August 2020).

Luke Henriques-Gomes "Robodebt class action: Coalition agrees to pay $1.2bn to settle lawsuit" *The Guardian* (online ed, United Kingdom, 16 Nov 2020).

Karen Hao "This is How Bias in AI Really Happens – And Why It's So Hard to Fix" *MIT Technology Review* (online ed, Cambridge, 14 February 2019).

Alex Hern "Why am I seeing this? New Facebook tool to demystify Newsfeed" *The Guardian* (online ed, United Kingdom, 1 April 2019).

Will Knight "The Dark Secret at the Heart of AI No One Really Knows How the Most Advanced Algorithms Do What They Do That Could Be a Problem." *MIT Technology Review* (online ed, Cambride, 11 April 2017).

Olivia Miwill "Malaysian Judiciary Makes History, Uses AI in Sentencing" *New Straits Times* (online ed, Malaysia, 19 February 2020).

Bernard Marr "Artificial Intelligence Has a Problem with Bias, Here's How to Tackle It" *Forbes* (online ed, United States of America, 19 January 2019).

Charles Towers-Clark "Can We Make Artificial Intelligence Accountable?" *Forbes* (online ed, United States of America, 19 September 2018).

J Websites

Liz Blythe, Joe Edwards and Rachel O'Brien "GDPR – A summary" (4 May 2018) Russell McVeagh <https://www.russellmcveagh.com/>.

George Coppolo "Criticisms of Federal Sentencing Guidelines" (16 July 1994) The Connecticut General Assembly <https://www.cga.ct.gov/PS94/rpt/olr/htm/94-R-0686.htm>.

Dr Rajiv Desai "Artificial Intelligence (AI)" (23 March 2017) Dr Rajiv Desai An Educational Blog <https://drrajivdesaimd.com/>.

Department of Human Services, Australian Government "Ticking All the Right Boxes" (4 September 2017) Australian Government Services Australia < https://www.servicesaustralia.gov.au/>.

Department of Immigration and Border Protection, Australian Government "Applying online or paper" Australian Government Department of Home Affairs < https://immi.homeaffairs.gov.au/>.

Noel Duan "When Beauty is in the Eye of the (Robo)Beholder" (20 March 2017) Arstechnica <https://arstechnica.com>.

"Explainable AI" simMachines <https://simmachines.com>.

Expert.ai Team "Explainable Artificial Intelligence Explained" (5 November 2020) expert.ai <https://www.expert.ai/?>.

European Commission "Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence" (21 April 2021) European Commission <https://ec.europa.eu/info/index_en>.

Will Knight "AI is Biased. Here's How Scientists Are Trying to Fix It" (19 December 2019) Wired <https://www.wired.com/>.

Ministry of Justice "Sentence indication" < https://www.justice.govt.nz/>.

Ministry of Justice "Official Information Act Requests" < https://www.justice.govt.nz/>.

New Zealand Government "New Algorithm Charter a world first" (28 July 2020) Beehive.govt.nz <https://www.beehive.govt.nz/>.

Privacy Commissioner "Privacy Act 2020" <https://privacy.org.nz/>.

Stats NZ "Algorithm charter for Aotearoa New Zealand" (20 November 2020) data.gov.nz <https://data.govt.nz>.

TC "What are algorithms?" (30 August 2017) The Economist <www.economist.com>.

Alexis Lee Watts "In Depth: Sentencing Guideline Grids" (11 January 2018) University of Minnesota <https://twin-cities.umn.edu/>.

James Vincent, "Amazon Reportedly Scraps Internal AI Recruiting Tool That Was Bias Against Women" (10 October 2018) The Verge <https://www.theverge.com/>.

John Zerilli, Alistair Knott, James Maclaurin, Colin Gavaghan "Transparency in Algorithmic and HumanDecision Making: Is there a double standard?" (5 September 2018) Philosophy and Technology <https://doi.org/10.1007/s13347-018-0330-6>.

K Letters and Submissions

Privacy Commissioner John Edwards "Submission to the Justice and Electoral Select Committee on the Privacy Bill 2017".

Associate Professer Gehan Gunasekara "Submission to the Justice and Electoral Select Committee on the Privacy Bill 2017".

L Interviews

Interview with AI Committee Members of Sabah and Sarawak Courts (the author, Zoom Meeting, 5 August 2021).