

https://dspace.org/

The use of predictive algorithm in judicial decision-making and the effects of 'machine bias' on human rights

Olorunju, Nokuthula

2019
https://hdl.handle.net/10539/29025

Downloaded from a DSpace repository. https://dspace.org/

NOKUTHULA OLORUNJU

THE USE OF PREDICTIVE ALGORITHMS IN JUDICIAL DECISION-MAKING AND THE EFFECTS OF 'MACHINE BIAS' ON HUMAN RIGHTS.*

Submitted in partial fulfilment of the requirements for the degree of

Masters of Laws by Coursework and Research Report

at the University of the Witwatersrand, Johannesburg.

* I am grateful my supervisor, Prof Cathi Albertyn for her generous assistance and encouragement through unchartered territories. I also extend immense gratitude to my parents for their support.

DECLARATION

I, Nokuthula Olorunju,

declare that this Research Report is my own unaided work. It is submitted in partial fulfillment of the requirements for the degree of Master of Laws (by Coursework and Research Report) at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

I have submitted my final Research Report through TurnItIn and have attached the report to my submission.

Word Count: 10579

SIGNATURE

531382
STUDENT NUMBER

11 June 2019
DATE

Abstract: *The Fourth Industrial revolution has seen an increase in technological advancements in adjudication. This research explores how judges are introduced to programs with predictive algorithms in efforts to enhance objectivity and efficiency. Many scholars are cautious about the role predictive, algorithmic programs can play in assisting judges to make decisions, especially when sentencing individuals. With the creation and use of these predictive algorithms come issues of transparency and accountability especially through the black box effect and machine bias. This is evident when predictive programs generate subjective results through hidden processes which lead to improper considerations of human right issues, as was seen in the case of State v. Loomis in the United States. Predictive programs lack the nuance and emotional intelligence, innate to human judges, to consider factors such as discretion, remorse and value judgements. The use of predictive programs will inevitably reach South African shores; the judiciary now has the opportunity to learn from the current international landscape by questioning the importance of predictive algorithms for decision-making. This research accepts that, in an effort to create efficiency, predictive algorithms may be inevitable in judicial decision-making, however, their use should be limited until better programs and systems of accountability are developed. The research ultimately proposes that enhanced transparency comes through ensuring programmers are held accountable, building judicial capacity and the creation of a symbiotic relationship of judicial oversight between judges and programmers of predictive algorithms.*

Table of Contents

Introduction 4

I Predictive Algorithms and Judicial Discretion 6

1.1. Understanding 'predictive algorithms' in context 6

1.2. The idea of judicial reasoning and adjudication 10

1.2.1. Independence and discretion of the judiciary 11

1.2.2. Example on the importance of discretion 13

1.3. The Judge vs. the machine 15

1.3.1. Judges interacting with predictive algorithms 15

1.3.2. The 'human element' 16

1.3.3. International judicial reactions to predictive algorithms 17

II 'Machine Bias and Human Rights 19

2.1. 'Machine Bias' and the Black Box Effect 19

2.1.1. 'Machine bias' 20

2.1.2. The black box effect 22

2.1.3. How predictive algorithms can be biased 24

2.2. COMPAS case study 26

2.2.1. Pitfalls of the case 28

2.2.2. Importance of the case 29

2.3. 'Machine Bias' vs. human rights 31

2.3.1. The cloak of 'due process' 31

2.3.2. The South African perspective 32

2.3.3. Court built by money 35

III Can predictive algorithms play a role in judicial decision-making? 35

3.1. Can 'machine bias' be counteracted? 35

3.1.1. Accountability of programmers 36

3.1.2. The "hybrid" judge: judicial symbiosis 39

CONCLUSION 42

BIBLIOGRAPHY 44

INTRODUCTION

The 4th Industrial Revolution[1] builds on the digital revolution and is characterised by the way technology has been embedded within societies, and blurs the lines between the physical and digital.[2] One sector facing this development is law, particularly in the introduction and use of predictive algorithms when making legal decisions. While the legal sector is often faulted for conservatism and a reluctance to join the digital revolution, there is nevertheless a steady progression towards paperless and efficient court processes.[3] Legal research is no longer exclusive to libraries and endless paging; there is a shift towards reliance on programs to compile vast amounts of information to assist the judiciary. This has led to the creation of algorithmic systems that have a predictive function aimed at assisting in predicting likely outcomes in cases.

While most case studies and literature on predictive algorithms in court processes are based in international jurisdictions like the United States of America (USA), these cases provide a window to what the future holds for judges globally. Although the idea of predictive algorithms is a combination of interdisciplinary concepts founded in law, computer science and information technology, attempts have been made to simplify these concepts. The focus of this research is on the socio-legal impact of algorithmic programs in the judiciary. The study examines the use of predictive algorithms in judicial decision-making and the potential implications on certain human rights. The paper also proposes that, due to

[1] Klaus Schwab 2016 'The fourth Industrial Revolution: what it means.' World Economic Forum. Available at: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond accessed June 2018

[2] Nicholas Davis 2016 'What is the fourth industrial revolution' World Economic Forum available at: https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/ accessed February 2018

[3] Pamela Katz 'Expert Robot: using artificial intelligence to assist judges in admitting scientific expert testimony' (2014) 24 *ALB. L.J. SCI. & TECH.* 1 at 28

evidence of a lack of accountability, programmers should be transparent and accountable when creating predictive algorithms.

The overall objectives of this study are to establish whether predictive algorithms have a role to play in the judiciary and whether the proposed efficiencies of predictive algorithms improve court processes. The study will also investigate whether algorithms will ultimately have a negative impact, through machine bias, on judicial decisions and human rights in terms of the Constitution of the Republic of South Africa.[4]

The research is addressed in three sections. Section 1 begins the discussion with a conceptual understanding of predictive algorithms; what informs judicial decision-making and the interaction between the two. The importance of judicial discretion is highlighted through the idea of the 'human element'. This is essential to fair decisions especially when certain circumstances require value judgements.

Section 2 considers how 'machine bias' can affect human rights. This is done by unpacking the concepts of 'machine bias' and the 'black box effect' and is illustrated through an international case study on recidivism where risk-assessment algorithms were investigated. An analysis is conducted on whether international methods of implementation of predictive programs, can be imitated in South Africa.

Section 3 addresses the accountability of algorithm programmers and proposes a symbiotic solution between the judiciary and use of predictive programs in relation to the previous sections. One example is of judicial oversight over predictive algorithms - limiting predictive programs exclusively to judicial support tools. Emphasis remains on accepting

[4] Act 108 of 1996

the inevitability of automated, predictive systems in courtrooms as part of the 4th Industrial Revolution.

The subject matter is a relatively recent culmination of two distinct fields, computer science and law, thus literature available is predominantly from international law journals, conference papers, periodicals and accredited online sources. This study has been conducted primarily through qualitative means whereby document analysis and literature review, involving established general principles of predictive algorithms, are used. The analysis includes a discussion on subject-matter literature, comparative analysis, academic case studies, national and foreign case law.

I. PREDICITIVE ALGORITHMS AND JUDICIAL DISCRETION

1.1. UNDERSTANDING 'PREDICTIVE ALGORITHMS' IN CONTEXT

Algorithms, commonly used in data processes, computer science and mathematics, are defined as a compilation of rules/instructions followed to solve a problem.[5] An algorithm works on the premise of a user giving an instruction, which a computer follows to give the fastest, most accurate result.[6]

Predictive algorithms are said to have statistics origins, specifically 'predictive analytics', and play an important role in the growth of big data and machine learning.[7] These algorithms use data, like personal information, to make predictions about an individual's

[5] Technopedia Dictionary 'Algorithm' available at: https://www.techopedia.com/definition/3739/algorithm accessed January 2019

[6] Andrea Roth 'Trial by Machine' (2016) 104 *Georgetown Law Journal* 1245 at 1269

[7] Dan Kopf. 'This is how computers "predict the future". Available at: https://qz.com/1261817/predictive-algorithms-are-not-all-that-complicated/ accessed January 2019

likely decisions and occur when technology, with maximum computational power, collects and manipulates vast amounts of data to generate more accurate and efficient results.[8] Essentially, the objective is to solve problems faster than humans do in various sectors.[9]

The mathematical aspect of an algorithm carries a quality of reassurance and logic, which people seem to immediately trust, because of an implied lack of emotion.[10] Currently, algorithms are utilised in most industries, determining credit applications,[11] job applications, insurance rates, consumer advertising and social network platforms.[12] Predictive algorithms have created a more efficient means of communication, research, time and case management. They can be considered the DNA of many daily processes in the Information Age.

In the legal field, the use of algorithms can be seen in many ways but two examples are highlighted: search tools and legal reasoning tools.

Search or 'data matching' tools use a combination of algorithms with the ability to sift through vast data amounts.[13] These algorithms compare different databases to find and match similar information, compiling results in ranks and in a manner best suited to the

---

[8] Danielle Keats Citron & Frank Pasquale 'The scored society: due proves for automated predictions.' (2014) 89 *Wash. Law Review* 1 at 3

[9] Kate Crawford & Jason Schultz 'Big data and due process: toward a framework to redress predictive algorithms, (2014) 55 *Boston Law Review* 93 at 96

[10] Will Knight 'Biased algorithms are everywhere and no one seems to care' July 2017 available at: https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/ accessed June 2018

[11] Citron et al *op cit* note 5 at 4, 8-18

[12] Tene O. & Polonetsky J. 'Taming the Golem: challenges of ethical algorithm decision making' (2017) 19 *North Carolina Journal of Law & Technology*, 125 at 133

[13] Dave Davies 'How search engine algorithms work: Everything you need to know'. Available from: https://www.searchenginejournal.com/how-search-algorithms-work/252301/ Accessed January 2019

given instruction.[14] Such programs would be classified as 'expert systems' because they use knowledge often exclusive to particular fields. Examples are search tools available on databases like Lexisnexis and Jutastats. These are not classified as predictive programs.[15]

Search tool algorithms also have the ability to 'fill in gaps' by providing more information or suggestions to users on related topics due to similarities found by the algorithm. An example is an algorithm that saves previous searches and suggests them at a later stage. These programs enjoy the advantage of transparency and consistency because the system provides users with the benefit of users understanding how a result was reached and functions as an advisory tool.[16] However, these systems have limitations in that they function solely on data inputted by creators when initially programmed. These programs cannot generate updated results without new data being input by users and are incapable of deviating from preset instructions.[17] Any new developments in the law still require human input/updates and the system requires consistent maintenance.[18]

Legal reasoning programs are a more recent form of predictive algorithms based on machine learning. Machine learning, which exists as a subset of AI, entails the ability of a program to learn automatically, without depending on human instruction and these programs are aimed at addressing the issue of machines depending on human users to evolve. Predictive algorithms have the potential to become self-evolutionary through

---

[14] Crawford op cit note 9 at 101

[15] Ibid

[16] Katz op cit note 3 at 33

[17] Robert J. Spagnoletti 'Using Artificial Intelligence to aid in the resolution of socioscientific disputes: a theory for the next generation' (1987) 2 *J.L & Tech* 101 at 102; Katz op cit note 3 at 30-32; Daniella Citron 'Technological due process (2008) 85 *Washington University LR* 1253 at 1260

[18] Ron Friedmann 'Automating legal advice: AI and expert systems' available at: https://prismlegal.com/automating-legal-advice-ai-and-expert-systems/ accessed November 2018

artificial intelligence.[19] Legal reasoning tools are also considered 'expert' systems and are already being used in USA, United Kingdom (UK), Canada, Australia and Brazil. A common example of a predictive program, used currently and deemed acceptable in courtrooms, is the e-discovery system.[20] This algorithm samples documents, identifies what is relevant and learns to detect the relevant from the irrelevant.[21] In law, attorneys use similar systems to analyse legal issues and compile possible arguments, generate opposing arguments and likely outcomes from compiled datasets to prepare for litigation.[22]

There are advantages to legal reasoning systems already in use in international courtrooms, especially in 'pro forma', small claim cases such as traffic offences or spousal maintenance cases in USA, UK and Brazil.[23] In traffic cases, predictive algorithms assess the minimum fine for an offence; any offences recorded against a driver and whether a licence is eligible for suspension. This lessens the pressure on a judge. Predictive algorithms are also used when parties qualify for legal aid and divorce settlements in Australia[24] and Canada[25] through a needs-based algorithmic assessment of financial records and the applicable marital regime, respectively.[26] However it was discovered, in 2015, that an algorithm/software used for divorce settlements in the UK had an error affecting thousands

[19] Spagnoletti op cit note 17 at 104; Citron et al op cit note 8 at 6
[20] Julie Sobowale "Beyond Imagination: How artificial intelligence is transforming the legal profession." (2016) *ABA Journal* 47 at 50
[21] Ibid
[22] Katz op cit note 3 at 30
[23] Julius Stone 'Man and machine in the search for justice.' (1964) 16 *Stanford Law Review* 515 at 559
[24] Katz op cit note 3 at 32
[25] Canadian Bar Association. 'Futures: transforming the delivery of legal services in Canada.' (2014) at 41
[26] Ibid

of divorce settlements by altering the financial status of spouses.[27] Spouses received higher settlements than they were entitled to.[28] This error arose when the development of a more online-friendly justice system to improve efficiency and finances was being encouraged.[29] The program creators claimed to be investigating the error, but one can presume they were indemnified from any liability, leaving the Justice Ministry to handle the fallout of the software. Thus, predictive algorithms may have the benefit of consistency; unfortunately this may sometimes be at the expense of equality, fairness and other human rights, as is discussed later.[30]

1.2. THE IDEA OF JUDICIAL REASONING AND ADJUDICATION

Adjudication is a complex 'machine' requiring various parts in order to function properly. It involves certain elements of emotional intelligence which encompasses discretion and value judgments, aspects unique to human beings - the 'human element' – and lacking in algorithms.[31] Much can be written about the concept, but the focus of the paper is to briefly discuss adjudication in a general context.

Although Roth approaches this discussion from a criminal justice point of view, she raises arguments that are still relevant to the important role of judges by stating:

---

[27] Owen Bowcott. 'Revealed: divorce software error hits thousands of settlements.' Dec 2015. Available from: https://www.theguardian.com/law/2015/dec/17/revealed-divorce-software-error-to-hit-thousands-of-settlements Accessed November 2018

[28] Ibid

[29] Ibid

[30] Crawford et al op cit note 9 at 101

[31] Citron op cit note 17 at 1304

'Still, there appears to be a wide agreement that human judges and juries have a role to play in softening mechanization's bluntness; that is, in acting as 'circuit breaker[s] in the State's machinery of justice'. Calls for robot judges and juries are typically met with derision, on grounds that machines are incapable of the individualised human judgement necessary to fully assess blameworthiness through a combination of complex factfinding, equitable discretion, and mercy.'[32]

1.2.1. Independence and discretion of the judiciary

Judicial authority is often established through founding laws such as a Constitution. The judiciary is considered independent and tasked with upholding the Constitution; all other organs of state are to assist the judiciary to uphold their independence, impartiality, accessibility and effectiveness.[33] The Judiciary is also tasked with protecting individuals' rights, so it is important that the judiciary is perceived as independent.[34] Judicial decision-making involves consistency and uniformity with elements of discretion and value judgments,[35] allowing an informed decision after weighing certain facts. Additionally, judicial reasoning, especially in South Africa, has played a significant role in interpretation, abolishment and limitations of laws with reference to constitutional rights and values.[36] The

[32] Roth op cit note 6 at 1247

[33] S165(2); s165(5)

[34] *South African Association of Personal Injury Lawyers v Heath* 2001 (1) SA 883 (CC) at [46]; K. O'Regan 'Checks and Balances reflections on the development of the doctrine of separation of powers under the South African constitution.' (2005) 8 *PER* 120 at 145

[35] Citron op cit note 17 at 1302

[36] F Du Bois '*Wille's Principles of South African Law*' 9 ed (2006) 65

South African judiciary has especially sought a substantive, constitutionally-inclined approach to legal reasoning.[37]

This is illustrated by Dworkin's principle of integrity, which states that a judge should be able to read and interpret any given text, with reference to applicable values but without attributing his/her personal convictions into their work.[38] Rights should be interpreted in the best possible light over time.[39] Judge Davis considers Dworkin's principles in a democratic South Africa, and states that judges have some measure of freedom/discretion when finding meaning in the text, but that does not mean that there is unfettered discretion. Discretion finds its limitations in precedent; therefore, judges cannot merely interpret law freely.[40]

Davis' argument draws more from Dworkin's theory of constructive interpretation which requires a judge to consider the law holistically to reach the best possible decision.[41] He mentions that a judge is constantly faced with his/her moral and political ideologies which can also reflect how s/he approaches a legal system.[42] When faced with a case, a judge can weigh either giving a judgment that aligns with those ideologies or, with more difficulty, one that is contrary to those ideologies to advance justice. However, the very idea of what constitutes justice often lies in the particular 'community's understanding of what justice demands', meaning that the idea of justice is interpretive and there can be no 'right' answer as Dworkin proposes.[43] A judge would have to shape the facts brought before him/her in a

---

[37] Ibid at 66

[38] Dennis Davis 'Dworkin: A viable theory of adjudication for the South African constitutional community?' 2004 *Acta Juridica* 96 at 97

[39] Ibid

[40] Ibid at 102

[41] A. Singh and M.Z. Bhero 'Judicial law-making: unlocking the creative powers of judges in terms of section 39(2) of the Constitution.' (2016) 19 *PER* 1 at 10

[42] Davis op cit note 38 at 103

[43] Ibid at 104-105

manner that advances justice according to the community's comprehension of what justice demands.[44]

South African judges have been granted a certain degree of discretion especially when interpreting the provisions of the Bill of Rights[45]- something reiterated by the Constitutional Court in *S v Makwanyane*.[46] This allows judges to retain the 'human element', especially when they have to consider exceptional circumstances.[47] In *S v Malgas*, the Court also emphasises that a court carrying out a sentence has discretion to deviate from a prescribed sentence, if rendering that sentence would be unjust in that it is disproportionate to the crime, the criminal and the needs of society.[48] This also highlights the importance of the judiciary maintaining individualised justice.[49]

In contrast to Dworkin's approach are legal positivists that, simply put, believed interpretation of law is a mechanical process where external, morality or discretionary considerations have no role in judicial decision-making.[50] Judges are expected to merely interpret the letter of the law to get to the intention of the legislature - leaving no room for exceptional circumstances. This was more evident in the apartheid regime and has much in common with the idea of predictive systems.[51]

1.2.2. Example of the importance of discretion

---

[44] Ibid at 105; Stone op cit note 23 at 553; Singh op cit note 41 at 10
[45] Singh op cit note 41 at 9
[46] 1995 (3) SA 391 (CC) at [9]
[47] Singh op cit note 41 at 9
[48] 2001 (2) SA 1222 (SCA) at [25]
[49] O'Regan op cit note 34 at 136
[50] Singh op cit note 41 at 5
[51] Ibid

Adjudicators should be guided by principles of equality, legal accountability, certainty, consistency and transparency when making decisions.[52] What becomes clear is that there is a responsibility on judges to justify their decisions. Their choices have to find meaning in that:

> 'Their responsibility requires acknowledgement that value-laden choices are made in the course of legal interpretation and that judges are responsible for the political, social and economic consequences of these choices, and should be judged accordingly. At the very least, a judge must acknowledge her political and moral responsibility in adjudication to justify the basis on which a decision was made, particularly in the interests of transparency and public deliberation.'[53]

A current example is the recent South African case of the Frankel 8. The Criminal Procedure Act[54] provides that sexual assault cases prescribe after a period of 20 years.[55] Following the consideration of the facts, the Constitutional Court was able to lift that 20 year prescription period understanding current social circumstances, weighing up extenuating factors and making a value judgment,.[56]

Should a predictive system have been in the same position, the likelihood of reaching the same decision would be unlikely, as all the data previously collected by the algorithm would be based exclusively on precedent. With precedent and a complete lack of discretion, an algorithm would categorically state that 20 years is the only acceptable time before prescription of such matters. Spagnoletti explains that decision-making is a process of

---

[52] Du Bois op cit note 36 at 80
[53] Davis op cit note 38 at 105-106
[54] Act 51 of 1977
[55] Section 18
[56] *Levenstein v Estate of the late Sidney Lewis Frankel* [2018] ZACC 16

factual evaluation AND value judgements[57] and these processes are not easily distinguishable.

1.3. THE JUDGE VS. THE MACHINE

The judiciary is no stranger to limited resources and thus any proposal towards a more efficient system is often greatly welcomed. Predictive systems can offer efficiency because fewer resources are spent on administration, paperwork and staff.[58] Judicial officials around the world, albeit unknowingly, already rely on predictive systems but the extent of such reliance is the point of contention. These are legal databases used to limit time spent on research and to focus more on interpretation and analysis of the facts.

1.3.1. Judges interacting with predictive algorithms

The interaction between the judiciary and predictive algorithms has been most evident in the USA. USA has the most recent records and case studies following predictive programs implemented in criminal court processes. These programs were initially exclusive to probation and parole departments to determine rehabilitation strategies for parolees.[59] They were later extended to the sentencing stage of criminal proceedings. This raised a red flag especially concerning the extent of reliance on predictive programs in adjudication and discussion will be focused on this issue. In South Africa, during sentencing, judges should be able to prioritise what issues are important: the 'triad of Zinn',[60] they should evaluate the severity of the crime, the offender's circumstances and public interest, exercising discretion

[57] Spagnoletti op cit note 17 at 107
[58] Citron op cit note 17 at 1259; Spagnoletti op cit note 17 at 114; Stone op cit note 23 at 549
[59] *State v Loomis* 881 N.W. 2d 749 (Wis 2016) [3]
[60] *S v Zinn* 1969(2) SA 537 (A) at 540;

when circumstances permit.[61] These factors are then weighed against the efficiency gains of algorithms and lessening the pressure on legal resources in using such systems.

1.3.2. The 'human element'

Judicial discretion relates to the idea of a 'human element' necessary for judges to make decisions - which predictive algorithms lack. Adjudication which exercises discretion, is rarely a 'black-and-white' or win/lose matter because different parties have different ideas of what constitutes winning a case.[62] Algorithms lack the ability to understand and relate on a nuanced, human level.[63] The 'human element' should be an important consideration, especially when making life-altering decisions which sometimes require considering exceptional elements and may result in making a value judgment rather than automated decisions.

This is not to say that judges are immune to biases, despite being considered the archetypes of impartiality in most societies due to the nature of their work.[64] Bail hearings and sentencing often fall to the discretion of a trial judge.[65] Research in USA has shown that a huge racial disparity exists where black Americans receive harsher sentences than their white counterparts do, without mechanical/predictive algorithms assisting the judge.[66] Bearing in mind that judges are not perfect, impartial beings, one must consider the use of predictive systems by judges, without proper parameters or oversight. Two things might occur: either judges use the machine to perpetuate an innate bias by giving harsher

---

[61] *S v Malgas* op cit note 48.
[62] Spagnolettti op cit note 17 at 116
[63] Ibid
[64] Willem H. Gravett 'The myth of objectivity: implicit racial bias and the law (Part 2)' (2017) 20 *PER* 1 at 9
[65] Ibid
[66] Ibid at 10-12

sentences and excusing their behaviour by blaming the predictive system; or they fear blame for relying on the system, choosing to 'overcorrect' bias by disregarding the system's recommendation and choosing a leaner sentence - which could be detrimental to punishment as a deterrent.[67] Effectively, this places judges in a precarious position, forced to create a delicate balance between monitoring both their discretion and not falling into either extreme. The balance may still end up being left to the judge's discretion - the human element - thus diminishing the assistive value of a predictive system.

1.3.3. International judicial reactions to predictive algorithms

Various international judiciaries, including Canada, European Union (EU) and the USA have engaged the subject of predictive algorithms and offer useful insight on the likely approach South African judges may adopt when faced with similar future concerns.

In 2011, the European Court of Justice (ECJ) considered the use of predictive algorithms on human rights grounds. The Court ruled that the use of gender in determining predictive insurance premiums was considered discrimination and therefore banned.[68] The determination of premiums operated on certain assumptions (although often said to be statistically based): that women live longer than men; that women are high risk employees due to child bearing; that men are breadwinners or that women are exclusively caregivers.[69] While to some extent this may be true, the fact that an algorithm made certain assumptions, based on what 'box you ticked', already raises alarms about the integrity and assessment of the software. If insurance companies can place individuals in certain categories, then they

---

[67] Ibid at 18

[68] *Association belge des Consommateurs Test-Achats ASBL & Others v Conseil des ministres* C-236/09

[69] Gendernet Practice Notes 'Employment, Labour, Social Affairs: Social insurance and equity between women and men.' Available from: www.oecd.dac/gender accessed September 2018

can charge customers for assumed characteristics that may not be accurate.[70] It also serves to perpetuate and reinforce issues of gender bias.[71] This also begs the question on whether this is just a loophole where blaming the software indemnifies creators from liability based on discrimination.

Additionally, the EU took a proactive step involving the oversight of automated decision-making. Article 22 of the EU's General Data Protection Regulation (GDPR),[72] which came into effect in 2018, bans solely automated decision-making which have legal or significant effects on an individual's life and, if such decisions are allowed, requires that they are subject to transparency and allow for human intervention to contest that decision.[73] While not expressly mentioned, it is implied that judicial decisions will be affected by this provision.

Canada followed the precedent set by the ECJ, but went a step further. The Ontario Human Rights Code forbids the insurance industry from using criteria involving race and religion, but allows for the use of age, sex, marital status and disability. The Code also allows complaints regarding discriminatory insurance practices to be lodged with the Human Rights Commission, making any discrimination an immediate constitutional matter.[74]

In *State v Loomis,* the U.S. Supreme Court had the opportunity to address the use of predictive programs by judges and acknowledged that technological tools used in adjudication continue to change. The Court accurately states that:

---

[70] Citron et al op cit note 5 at 4; 8-18
[71] Ibid
[72] EU Regulation (2016/679)
[73] Guido Noto La Diega 'Against the dehumanisation of decision-making.' (2018) 9 *JIPITEC* 3 at 5
[74] Thomas Flanagan 'Insurance, Human Rights and Equality Rights in Canada: When is discrimination reasonable?' (1985) 18 *Canadian J. of Political Science* 715 at 716, 724

> 'The concerns we address today may very well be alleviated in the future. It is incumbent upon the criminal justice system to recognise that in the coming months and years, additional research data will become available... As data changes, our use of evidence-based tools will have to change as well. The justice system must keep up with the research and continuously assess the use of these tools.'[75]

The importance of algorithms to judicial decision-making is valid and can have life-altering consequences. Roth uses the term 'automation complacency'[76] to describe when a human slowly defers to the judgement of a machine, begins to forget the importance of discretion, and fails to notice exceptional or extenuating factors in cases. This results in potentially reducing each individual who appears before the court to a statistic. This is not to say that judges cannot rely on predictive algorithms as a means to be more efficient; but caution should be taken on the extent of reliance. While a mechanical approach to adjudication carries the appeal of results at face value, it can also compromise important values protected through human discretion.[77]

II. 'MACHINE BIAS' AND HUMAN RIGHTS

2.1. 'MACHINE BIAS' AND THE BLACK BOX EFFECT

Predictive algorithms, like all systems, are not fail-safe. They are faulted for producing biased outcomes. These outcomes are usually attributed to 'machine bias' and the black box

[75] *Loomis* supra note 59 para 4
[76] Roth op cit note 6 at 1279
[77] Ibid at 1249

effect. These terms are discussed from a socio-legal perspective in this research and should not be mistaken with terms in the scientific context.

2.1.1. 'Machine bias'

'Machine bias' is a concept that is still in early stages of research and investigation, but has garnered attention for how long it has occurred without raising red flags.[78] It should be stated from the onset that a system, of its own accord, does not have the ability to be biased in the socio-legal sense because it is the sum total of data collection, processing and distribution.[79] A machine only processes data and, thus far, has neither the emotional intelligence nor the ability to reflect nuances distinct to humans. Its use in courtrooms is largely motivated by the need for uniformity and rational punishment.[80]

Additionally, to objectively state that an algorithm is biased, one must have created objective boundaries of what constitutes ethical standards/rules regarding discrimination/bias.[81] That, alone, is a highly contentious issue. [82] If those boundaries do not even exist in terms of human reasoning, then an algorithm cannot realistically be expected to produce results that are perfectly bias-free since it is created by humans with biases.[83]

'Machine bias', in this context, refers to bias fundamentally linked to the algorithm created by the programmer of the system. Computer scientists argue that machines cannot be

---

[78] Knight op cit note 10
[79] Matthew T. Nowacheck 'Why robots can't become racist and, why humans can.' (2014) 9 *Phaenex Journal* 57 at 67.
[80] Roth op cit note 6 at 1267
[81] Tene et al op cit note 12 at 166
[82] Ibid
[83] Ibid

biased; however, humans create systems which can carry the characteristics of their subconscious biases.[84] An algorithm can also show bias because of collaboration between programmers giving conflicting instructions.[85] An example is a system created between computer scientists and legal experts.[86] The legal expert analyses certain facts, using laws and rules, guiding the engineer through the process in order for the system to learn to imitate legal analysis and operate at a more efficient level.[87]

Thus unlike the scientific field wherein formulae and methods are often set in stone for long periods, the legal field often has a more fluid and dynamic approach which can result in several possible conclusions. In other words, the system becomes potentially biased towards the particular legal expert's reasoning and there is no guarantee that such reasoning is the only way to analyse the given data. Therefore, every time the same data entry is made, the machine produces a result that mimic's the legal expert's thinking - imputing human biases and 'insecurities' into factual situations.[88] This can result in inefficiencies because experts reason differently even if they reach the same conclusion. A system that mimics the bias of the creators could potentially taint the outcome of its results.

Machine bias can also appear through inconsistencies of biased datasets collected.[89] A machine essentially attributes certain characteristics to a new set of data because of the similarities the new data shares with old datasets. The parameters of algorithms are rarely,

[84] Crawford et al op cit note 9 at 117, 120
[85] Citron op cit note 17 at 1260
[86] Spagnoletti op cit note 17 at 109
[87] Ibid
[88] Tene et al op cit note 12 at 155;
[89] Ibid at 136

if ever, neutral.[90] Scholars have been at the forefront of identifying such machine bias in credit score ratings, commercial risk-management software, voting software, criminal detection systems and even in legal research systems.[91]

2.1.2 The black box effect

A black box is a tool often associated with a data-monitoring and collection device used predominantly in aeronautics.[92] The 'black box effect' however, refers to a program whereby one can only monitor data inputs and outcomes but is not privy to the system's internal processing and analysis.[93] It is characterised by the fact that its inner workings are opaque. Simply put, the system mimics the human brain: the eyes see light and yet the brain analyses and differentiates light into colours - the brain does not explain to the eyes why green is green or why red is red. While, over the years, brain-mapping has been able to explain how the brain works, a human being identifying colour is still an automatic or 'opaque' process and the black box effect mimics that process.

The black box effect finds relevance especially in the Age of Big Data where massive amounts of data is collected and stored.[94] It has negative effects on objective and fair adjudication because there is no clarity on the algorithm's analysis or verification on the system's decision-making process.[95] This makes any form of appeal or review impossible

---

[90] Ibid

[91] Roth op cit note 7 at 1270

[92] National Geographic 'What is a black box?' available from: https://www.nationalgeographic.com.au/engineering/what-is-a-black-box.aspx accessed January 2019

[93] Frank Pasquale 'The Black Box Society: the secret algorithms that control money and information' (2015) 6

[94] Crawford op cit note 9 at 96

[95] Roth op cit note 6 at 1269-1273

because there are no details on what precedents or relevant factors were used in its results; or whether the system relied on a different jurisdiction.[96] The algorithms' obscurity and complexity can perpetuate unaccountability and could potentially render an incorrect prediction/decision due to incorrect data analysis.[97] Thus even if proven that predictive algorithms are efficient, results from obscure processes cannot find purchase in a legal sector built on principles of transparency, accountability and equality. Pasquale affirms: 'Faulty data, invalid assumptions, and defective models can't be corrected when they are hidden.'[98]

Tene and Plonetsky build on the above by discussing the idea of how the 'black box' effect began with the human mind. They argue that humans are just as, if not more, susceptible to bias, discrimination and prejudicial thinking than machines.[99] However we must not discount the fact that humans can still reason out and reflect over their decisions.[100] As algorithms are created to reflect real-world circumstances, not utopian results, it makes sense that something like the black box effect exists because, as mentioned before, bias subconsciously exists in all humans and without the human element of change and/or self-correction, the system merely reflects what already exists in a society.[101]

Another important aspect is that, although algorithms aim to combat human errors, the result, more often than not, is the creation of new, unforeseen errors like machine bias and

[96] Citron op cit note 17 at 1253
[97] Pasquale op cit note 93 at 18, 217
[98] Ibid at 21
[99] Tene et al op cit note 12 at 162
[100] Stone op cit note 23 at 555-556
[101] Tene et al op cit note 12 at 164

the black box effect in adjudication.[102] While these concepts are disadvantageous effects of predictive systems, one must still question whether they warrant complete disregard by the judiciary, and if not, what kind of algorithms can be used by the judiciary and in what circumstances are they acceptable tools. This is best understood through case studies.

2.1.3. How predictive algorithms can be biased

The idea behind a biased algorithm is that the algorithm only reflects what general trends currently exist based on a limited sample of data.[103] Some argue that altering algorithms to only show objective results imputes an incorrect reflection of society which is inherently biased.[104] One must still question whether systems that perpetuate bias should be approved for use by judges, or should only be allowed into the market when the data collected accurately reflects a more 'objective view'. This also begs the questions: first, can objectivity ever be perfect in an imperfect society? Secondly, does the idea of objectivity and accuracy just mean current social/cultural biases?[105] If objectivity is a moving target, a relative idea, then algorithms would be expected to adapt to different accepted standards of objectivity, based on the relevant society, which creates added complications.

When considering the above, one such complication is that the development and implementation of these systems can occur in very distinct geographic locations. A system developed in America or Europe would be quite difficult to implement in Africa. While

---

[102] Katherine Freeman 'Algorithmic injustice: how the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis' (2016) 18 *N.C. Jour. of law & Tech* 75 at 84
[103] Tene et al op cit note 12 at 153
[104] Ibid at 135; 153
[105] Tene et al op cit note 12 at 153

there are valid arguments that systems can be tailored to suit the African legal framework, one must also consider that these systems are not necessarily created with a diverse cultural background, but are more "Euro-centric" in nature. For example, how does one deal with practices of 'Umhlanga' (virginity testing) in the Zulu tradition,[106] or traditional male circumcision practices in various African cultures, or the West African 'Sharo/Shadi' practice of the Fulani tribe whereby young men are flogged in order to prove manhood and the faces of young girls are tattooed as a sign of womanhood?[107] From a purely human perspective, there are opposing views about these practices and some would argue they border on abusive practices. However, without the requisite knowledge and understanding of diverse cultural backgrounds, it would be easy to mechanically attribute criminality to certain accepted practices. In other words, an algorithm created by an American or European may recommend a practice is child abuse to a judge, when an African might identify and acknowledge it as an accepted practice.

Consequently, merely adjusting an algorithm not to produce biased outcomes does not necessarily solve the problem; it merely fuels existing attitudes of ignorance or misconception. We should be careful not to create a 'utopian' experience that does not appropriately reflect the world today. While it is also argued that some bias is necessary and may even be negligible, such arguments can only be acceptable when the process to reach decisions are clear, concise and capable of explanation - only when there is no opacity. However, with the growing obscurity of current technologies, like the black box

---

[106] Louise Vincent 'Virginity testing in South Africa: Re-traditioning the postcolony' (2006) 8 *Culture health & Sexuality* 18 at 18-19

[107] National Geographic.'Fulani Initiation Rites' October 2017, available at: https://video.nationalgeographic.com/video/benin_fulaniinitiation accessed June 2018

effect, and no clear parameters, one can never be reasonably sure that the resulting biases are negligible or even necessary.

2.2. COMPAS CASE STUDY

In 2014, investigative journalists working for the independent, non-profit newsroom, ProPublica, conducted an investigation into risk assessment algorithms used by judges, during the sentencing stage, in the United States Department of Justice.[108] The investigation was to assess how predictive algorithms affect issues of race and equality in the criminal justice system, while attempting to evoke social consciousness.[109] This particular case study is relevant as a rare example of judges using predictive algorithms practically, as most systems proposed for judicial use usually remain prototypes.

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) program was created by a private company, Equivalent (formerly Northpointe), to predict the risk of previously prosecuted individuals committing future crimes. This was done by assigning scores to defendants and providing those scores to judges during sentencing.[110] The predominant issue arising from this investigation was that individuals of African American and Hispanic backgrounds received scores predicting a higher rate of re-offending or criminal violence than those of Caucasian backgrounds. Therefore, those

---

[108]ProPublica 'About' available at: https://www.propublica.org/about/ accessed February 2018

[109] Angwin J. Larson J, Mattu S. & Kirchner L. 'Machine Bias' May 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing accessed February 2018; Larson J. Mattu S. Kirchner L. & Angwin J. 'How we analysed the COMPAS Recidivism algorithm' May 2016. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Accessed February 2018

[110] Ed Yong 'A popular algorithm is no better at predicting crimes than random people' 2018 available at: https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/ accessed June 2018

defendants were labelled as 'high risk' and subjected to longer imprisonment.[111] When asked to explain how this algorithm worked, the company claimed that the proprietary nature of the algorithm meant that it could not be disclosed to defendants or the public for investigation into exactly how the system reached its results.[112]

A fascinating aspect involves the data collected by the COMPAS algorithm. Race was not an explicit question, rather this algorithm asked alleged offenders questions relating to their personal beliefs, personal history, trust levels and thought processes in certain scenarios.[113] The program took into account: personality traits, presence of certain characteristics such as anger, education, vocation; social background and environment, social isolation, substance abuse and peer relations.[114] These factors were rated similarly to that of a questionnaire; on a scale from strongly disagree to strongly agree. The COMPAS algorithm apparently also made adjustments for deceptive answers, random answers (answering carelessly) as well as inconsistency in answering.[115] This meant that should the alleged offender not pay attention to how s/he answered the questions, it worked against them. The gathered data was then processed through the algorithm which provided a score and classified the offender as either 'low, medium or high' risk with a likelihood to re-offend.[116]

The implications of such a program raises both the black box effect, by not disclosing to the defendant how the program works, and machine bias, stemming from the instructions the creators/engineers have programmed into the system. The program perpetuates the idea that

[111] Angwin et al op cit note 109
[112] Ibid
[113] Freeman op cit note 102 at 76
[114] Ibid
[115] Ibid at 77
[116] Ibid

backgrounds, beliefs, current emotions and perhaps even unforeseen circumstances, exclude certain individuals from fair trial and rehabilitation based on an obscure estimation that they would re-offend.

The COMPAS case was brought before the Wisconsin State Supreme Court in *State v Loomis.*[117].This case is the first to address the legality of using predictive algorithms during sentencing. The defendant, Eric Loomis, raised three issues before the court. Firstly, a violation in using a defendant's information in sentencing without defendants verifying the accuracy of the information collected; secondly, violating the right to an individualised sentence; thirdly, the algorithm's improper use of stereotypes during assessments.[118] These issues meant the accused was unable to prepare a proper defence because he had no access to how the collected data was rated, if any bias was evident or what factors were taken into account. Simply put, it is difficult to formulate an argument against something unknown.

2.2.1. Pitfalls of the case

The Wisconsin Supreme Court decided against Loomis. This judgement is criticised for its failure to protect the rights of individuals and it is argued that the Court had no clear understanding of the algorithm.[119] The Court reasoned that a defendant had to read the report generated by the COMPAS programme to refute the results of the algorithm.[120] The

[117] Loomis supra note 59
[118] Ibid para 34
[119] Ibid at 82
[120] *Loomis* supra para 52-58

report, however, failed to provide an explanation for the algorithm's reasoning in its decision-making. For example, it did not explain the weighting of the factors considered.[121]

The court's reasoning implies that individuals appearing before a court using a predictive algorithm should have some understanding of how the algorithm works - which is unreasonable. Although the court acknowledged the importance of having complete and accurate information during sentencing and that algorithms are only assistive tools, the court itself still battled with taking a decisive stance on whether or not predictive algorithms should be used by the judiciary.[122] The court conceded that the program had discrepancies regarding racial issues,[123] and was based on predicting group data rather than individual data.[124]

Other criticisms included the fact that the creating company, Northpointe, already had several million dollar contracts with the State and auditing of their systems were carried out by affiliates of the company - meaning a lack of independent investigations into systems.[125] This is particularly interesting when one considers that the motivation behind bias is generally the element of self-interest.

2.2.2. Importance of the case

Despite the criticisms of the *Loomis* judgement, the Wisconsin Supreme Court made several important points about predictive programs. Relying on predictive programs, as

[121] Ibid para 66; Freeman op cit note 102 at 82
[122] Supra para 38-40
[123] Ibid para 66
[124] Ibid para 68
[125] Freeman op cit note 102 at 83

determinative factors, for sentencing would be contrary to individualised justice.[126] Judges using predictive programs should always practice discretion and these programs cannot be used to determine the length and severity of the sentence.[127] Furthermore, judges must be cautious of their dependence on such programs as they can threaten the rights of individuals. The Court warned other courts to properly assess the accuracy of algorithms going forward.

In response to ProPublica, Flores et al argue that algorithms, like COMPAS, used to inform decisions relating to sentencing, have proven to predict outcomes more accurately than professional human judgment - which lacks structure.[128] In other words predictive programs are better judges of character than actual judges are, thus implying that actuarial-based algorithms are superior to judicial reasoning. They argue further that the use of predictive programs is a result of years of research.[129] Their argument, although clearly advocating for more structure, seems to support a mechanical approach to decision-making with accuracy as the determining factor, rather than individual analysis.

ProPublica's approach to the COMPAS case adopted a sociological perspective, and perhaps a more legally aligned approach, which considered the questionable fact that algorithms are being used in life-altering decisions, a volatile legal and social issue. Flores et al's approach is centred on the science of predictive algorithms.[130] These arguments are

---

[126] Supra para 68-69

[127] Ibid para 95-97

[128] Flores, Anthony; Bechtel, Kristin & Lowenkamp, Christopher. 'False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks.' (2016) 80 *Federal probation* 14;

[129] Ibid

[130] Ibid at 6

prime examples of how consensus can be difficult even as technology blurs the lines between computational and social science.

2.3. MACHINE BIAS vs. HUMAN RIGHTS

The proposal that judges be replaced by machines dates back to the 1970s.[131] The proposal has been met with scepticism; with most agreeing that it would be far better to be tried by a human than a machine because the beginning of adjudication by machines is also the beginning of society unravelling.[132] Courts have yet to fully engage on and interpret how predictive algorithms interact with human rights and all discussions thus far are largely academic.

2.3.1. The cloak of 'due process'

The biggest debate surrounding predictive algorithms in judicial decisions and human rights is centred on the concept of 'due process' under the Fourteenth Amendment to the United States Constitution. The Fourteenth Amendment states: 'No person... shall be deprived of life, liberty, or property, without due process of law'.[133] 'Due process' relates to fairness, equality, transparency, dignity and the importance of a fair hearing before an impartial adjudicator.[134] It embodies both the substantive and procedural aspects of law. The corresponding rights to 'due process' within the South African Constitution are

[131] Roth op cit note 7 at 1285-1287
[132] Ibid
[133] US Constitution, Amendment 14
[134] Crawford op cit note 9 at 112; 119

contained in Chapter 2 of the Constitution and include the rights to equality, access to information and the right to a fair hearing.[135]

Therefore, due process is a holistic collection of values that involve protecting both the substance of human rights and procedural rights. It places a responsibility on judges to ensure transparent and accountable processes are evident, even when predictive systems are used. Examples include a judge who considers 'due process' through the lens of discrimination substantively (e.g. should a woman be granted a lesser sentence for sexually assaulting a man because of an assumed bias that men are stronger than women?); or procedurally (e.g. was a defendant denied the opportunity to cross-examine state witnesses?).

2.3.2. The South African perspective

South Africa has taken great steps to overcome the injustices of the past but the fact remains that current demographics still reflect many social disparities surrounding human rights issues of race, religion and gender, which also exist as implicit biases.[136] Understanding human rights is a delicate process that still requires finesse during interpretation, especially by judges. It may be easier for judges to err on the side of caution and limit interactions with predictive programs to procedural aspects as they might only ever be seen as a mechanical reflection of society's imperfection.

On the other hand, it has been proposed that the use of technology will make access to justice more feasible for the common person, which will optimise the entire justice

[135] Sections 9; 32; 34
[136] Gravett op cit note 64 at 12

system.[137] The idea of justice can just as easily be destroyed if individuals believe that machines determine their fate. When the extent of a judge's reliance on an algorithm outweighs his/her ability to understand how the program reached a decision, or a judge fails adequately to question its legal reasoning, then the protection of human rights has taken a secondary role to a machine and the judge fails to uphold the principles of due process.

A foreseeable problem is the capacity of South African judges to balance the protection of human rights with the regulation of predictive programs without falling into abovementioned extremes.[138] This refers to judges using predictive programs either to justify their implicit biases or to 'overcorrect' bias by disregarding valid factors raised by the program for fear of appearing too harsh or lenient, or being racist or sexist. A practical example includes an accused being a black woman, a subconscious thought might exist either that her race/background justifies a violent nature, or that her gender makes her incapable of committing violent crimes.

Individuals with human rights reasonably expect not to be deprived of those rights. This means that they should have the right to be heard and to challenge incorrect data considered by programs.[139] Algorithms can be subject to typographical errors, bugs or mistranslations and these errors can often go undetected for years. When programs are making important decisions that could violate the human rights, this could end up costing the State millions in compensation claims.[140]

---

[137] B. Alarie, A. Niblett & A. Yoon 'Law in the Future' (2017) 3 *Univ. of Toronto LJ* (Focus Feature)
[138] Refer to point 1.3.2. above
[139] Crawford op cit note 9 at 127
[140] Roth op cit note 6 at 1275-1276

Therefore, at this particular moment, human rights should always be able to outweigh the idea of automation and 'machine bias'. Flanagan validly states that human rights are based on the premise that people should not be differentiated based on 'immutable characteristics'.[141] When applied to the idea of using predictive algorithms, a person's sex or race should not affect an algorithm's calculation as to whether they are likely to commit or repeat an offence once convicted. What is paramount is that justice is deemed served based on the convictions of the community, as is argued by Davis.[142] This again highlights the importance of human discretion in decision-making and individualised 'due process'.

The idea of exposing judges to algorithms in South Africa is not a new one. Gravett discusses the example of using algorithms in relation to the selection of judges by the Judicial Services Commission (JSC) - which has been faulted for inconsistency in its selection processes. He points out that when people compete with machines, 'our sympathies lie with our fellow human. The aversion to algorithms that make decisions that affect humans is rooted in the strong preference that many people have for the natural over the synthetic or artificial.'[143] However, he recommends that there should be some effort towards more structure and consistency in choosing judges than what he sees as the mere 'gut intuition' of JSC members.[144] This is an interesting consideration because unsuccessful judicial candidates could potentially institute claims of 'automated discrimination' if they are unhappy with the JSC's selection process.

---

[141] Flanagan op cit note 74 at 733

[142] Davis op cit note 38

[143] Willem Gravett 'Towards an algorithmic model of judicial appointment: the necessity for radical revision of the Judicial Service Commission's interview procedures' (2017) 80 *THRHR* 267 at 284

[144] Ibid

2.3.3. Court built by money

While technological innovations are predominantly from the private sector and it would defeat the advancement of technology to stifle them; one must proceed with caution to ensure that the protection of human rights is not defeated by the drive to create wealth. The private sector is often at the helm of developing systems used by the public sector, including courtrooms, and private stakeholders are driven by a profit maximising perspective. Furthermore, because similar systems are prevalent in other commercial and government sectors, the legal sector is definitely not exempt from predictive technology. Society, however, should be cautious not to simply consent to or rubberstamp private sector creations without proper assessments. The private sector cannot become an indirect final arbiter of legal developments that carry vast consequences.[145]

These systems may have been created with the best intentions, as efficiency and accuracy are a necessity in courtrooms today. But these potential benefits cannot overshadow the need to ensure the protection of individualism and human rights.[146] We should learn from the COMPAS experience.

III. ROLE OF PREDICTIVE ALGORITHMS IN JUDICIAL DECISIONS

3.1. CAN 'MACHINE BIAS' BE COUNTERACTED?

In every situation, society must decide whether the potential benefits are worth the associated risks. Different persons will have conflicting views as to the risks that justify a

[145] Tene op cit note 12 at 141
[146] Freeman op cit note 102 at 79

particular benefit.[147] This section discusses solutions through the idea of programmers' accountability and a judicial hybrid system for effective use of a predictive system.

3.1.1. Accountability of programmers

No algorithm is immune to its creator's values,[148] they are the 'brainchild' of humans and trained on human generated data with inherent biases.[149] When creating predictive programs for use by the judiciary, programmers are inadvertently playing a role in rule/law creation by translating laws from human language to computer code.[150] The transition from law into an algorithm can indirectly contain bias stemming from a programmer's 'limited' understanding or interpretation of the law.[151] Although programmers may not have the intention to change or distort the law, it does not change the fact that the translation can affect how the law is interpreted.[152] Therefore algorithm creators/programmers should be expected to follow ethical processes and auditing standards and rules.[153]

The accountability of programmers begins the moment transparency of the algorithm is lost (falls into a black box effect freefall). Even the consultation of experts in creating predictive programs is not a remedy for replacing the discretion involved in judicial decision-making, as advocated by Citron.[154] When one considers the fact that even the legislature and judiciary follow rules and standards when creating and interpreting laws, a similar standard should be followed by programmers.

---

[147] Spagnoletti op cit note 17 at 101
[148] Tene op cit note 12 at 128;
[149] Ibid at 130
[150] Citron op cit note 17 at 1254
[151] Ibid at 1262
[152] Ibid at 1254
[153] Tene op cit note 12 at 141
[154] Citron op cit note 14 at 1255, 1261

Programmers should contribute towards transparency and accountability in their programs to avoid suspicions of human engineering or bias by taking responsibility when gaps are found in the programs. They can do this by disclosing that the algorithm is tailored to ethical norms or values.[155] Programmers rely on the argument that source codes used to create algorithms are trade secrets, meaning those who wish to challenge their accuracy are not given access. This also means that these systems are not subject to appropriate, independent reviews. Courts are also reluctant to compel the programmers to disclose the information.[156]

Freeman proposes a relevant solution to 'machine bias' resultant from the COMPAS algorithm, namely, the conversion of the algorithm from proprietary to open-source.[157] This enables potential defendants to inquire, and seek counsel, as to the information used in the algorithm. Open source algorithms enhance the ability of relevant parties to question, point out weaknesses and evaluate the accuracy and consistency of any algorithm. It also forces creators to work harder to build better systems.[158]

Other scholars divert from Freeman's approach, suggesting that responsibility does not have to come with revealing proprietary source codes but rather through public accountability. The writer believes that the limited disclosure of source codes is necessary, especially if programs are made subject to rigorous tests by experts in the sciences and law until there is unequivocal proof that the algorithm meets acceptable standards. Open source disclosures will compel creators tread carefully and be transparent before program

[155] Tene op cit note 12 at 132
[156] Roth op cit 6 at 1274-1275
[157] Freeman op cit 102 at 85
[158] Ibid

proposals are brought before the judiciary. In terms of protecting the proprietary nature of the code, programmers can use non-disclosure agreements to protect their work. Legal analysis and decision-making require a transparency that should be mandatory and the same should apply to programmers.

In terms of accountability, Pasquale emphasises that success in accountability requires human judgement.[159] Once again, it is vital to ensure that in the move towards an automated society, prejudice and discrimination do not become intrinsic to the 'codes' we live by.[160] If independent computer science and legal experts can judge the value of a program, through intensive assessments, and jointly agree that a program should be used, then programmers are a step closer to accountability.[161]

In contrast, accountability being attributed to programmers can also negatively affect innovation, as they do not want to be responsible for life-altering decisions because of an algorithm. To clarify, by accountability, the writer is not proposing that programmers be sued for damages or suffer criminal charges per se; rather that code disclosure becomes a requirement for creators. This can be done through independent auditing studies, tests and, most importantly, scrutiny by the judiciary. When a program is created to advocate transparency and is explicit as to its limitations, it is more likely to gain acceptance by clients in the relevant market and it counteracts the presumption that machines are infallible.[162]

---

[159] Pasquale op cit 93 at 213
[160] Ibid
[161] Tene op cit note 12 at 171
[162] Citron op cit note 17 at 1304

3.1.2. The man and machine hybrid: judicial symbiosis

The idea of a 'hybrid' system speaks to judicial oversight regarding predictive algorithms. It means creating a symbiotic relationship between judges and predictive programs, but with judges having the final say. Judges interviewed as part of the ProPublica investigation were of the view that, while the predictive systems provided a concise summary of the defendants' histories, some judges still relied on their personal expertise and discretion in making appropriate decisions.[163] Therefore most judges said they merely used the programs as a consultant or an assistive tool.

Before predictive programs get into courtrooms, judges should have the opportunity to question the logic behind the processed data. Systems should come with the disclaimer that reliance on the algorithm is at judges' exclusive discretion and they should have the ability to consult knowledgeable experts regarding systems. Companies that spearhead these systems and the government departments that endorse them are not necessarily motivated towards monitoring and regulation,[164] it may be left to the judiciary to limit the extent of implementation and encourage scrutiny.

Concerns exist regarding judicial capacity to operate predictive programs and the added pressure to their workload. While training judges to use predictive programs means greater responsibility for the judiciary, it ensures that judges have a complete understanding of the program and can identify when the algorithm is contrary to the objectives, rights and values of the Constitution. This is not to say that predictive programs are not advantageous, they can be significant when the letter of the law (as with the traffic offences) rather than

[163] Angwin et al op cit 109
[164] Knight op cit note 10

discretion is required; there just cannot be blind reliance on any system regardless of alterations or 'fixes'.[165]

What also becomes clear is that the judiciary has a role to play in the practical significance of how and when predictive algorithms can and should be used in adjudication matters. One practical example is that the Office of the Chief Justice (OCJ) can spearhead capacity building through 'outsourcing' judicial training to companies with the right expertise in predictive programs. The OCJ can receive detailed reports from independent expert advisors or a regulatory body tasked with assessing the transparency of programs, following which a sample size of judges will conduct assessments and provide feedback on the programs.

Another example is in relation to the rules on granting leave to appeal[166] and the idea of giving individuals the right to appeal or review automated decisions.[167] Perhaps a defendant can make a substantially researched and reasoned application against the use of an algorithm in their matter. Some may also argue that predictive algorithms speak more to procedural aspects and are more suited to a pre-trial or even review.[168] There are existing mechanisms possibly to regulate the implementation of predictive algorithms in the South African judiciary to counteract potential bias - at least for those that will be introduced in the coming years.

Additionally, before implementation of predictive systems can occur, there has to be strictly monitored, testing phases and an established regulatory body consisting of computer

[165] Citron op cit note 17 at 1303
[166] Rules 49-52 Uniform Rules of Court (As amended) in GN R315 GG 19834 of 12 March 1999
[167] Citron op cit note 17 above at 1306
[168] Rules 37; 53

science, information technology, psychology and legal experts. These phases are to pinpoint areas of bias, misinterpretation of law converted to code, obscurity, distortion of laws and uncertainty.[169] They are also to create a baseline from which ethical standards and protocols exist for creators who wish to make system creation bids in the judiciary.

Some recommendations for a regulatory body include the establishment of a neutral arbiter to investigate allegations of bias similar to the role of the Information Regulator established by the Protection of Personal Information Act.[170] This independent body will audit and monitor the digital footprint of systems and can adequately address concerns raised by the judiciary. The problem with this proposal is that it encroaches on the independence of the judiciary and can create unforeseeable problems, unless the arbiter's function is limited to testing programs and assisting the OCJ in training for judges.

For all the disadvantages contemplated by the introduction of these predictive algorithmic programs, one cannot deny that their introduction will be the catalyst that forces the judiciary to break into the 4th Industrial Revolution.[171] If judges are trained properly and gain accurate understanding of these systems, it would revolutionise legal processes. Examples include programs that categorise matters on a judge's case list and give the judge an estimate on the duration of the matters, or programs that flag international cases applicable to a judge's matter. It could mean less work for support staff and proper time management.

[169] Citron op cit note 17 at 1311
[170] Section 41 of Act 4 of 2013; Crawford op cit 9 at 127
[171] Freeman op cit 102 at 86

It is important to understand that rather than blame algorithms, one should address the laws and norms that currently fail to protect individuals from social inequities. Much like the Ontario Code, where existing law is used to address the possible effects of biased systems in insurance, the focus should be on strengthening existing laws to address unforeseen issues arising from the use of predictive algorithms.

The fact remains that there will always be a need for human adjudicators as they have the advantages of providing express analysis and the opportunity to challenge their decisions reasonably. They offer the 'human element' still necessary to understand the complexity of society. Notwithstanding that any implementation is bound to be contested by the public interest organisations, in an effort to protect the rights of South Africans; it is a necessary 'evil' to face the reality that predictive algorithms are the expected next step towards the 4th Industrial Revolution.

CONCLUSION

The study into the role of predictive algorithms in the judiciary is still new and complicated, which can lead to a labyrinth of questions requiring further research. This study has shown that for all its imperfections, predictive algorithms still have a role in the judiciary and that role will increase over time with technological developments. The study also analysed the dynamics that currently exist between the idea of 'machine bias', the black box effect and judicial decision-making with the potential effect that they would have on certain human rights. There is a need for a process that combines the humanity of judges with the efficiencies of predictive algorithms.

It is the writer's view, subject to empirical verification, that the South African judiciary be made aware of the changing tides of decision-making and learns from the various international discussions that have come about from the use of predictive algorithms and similar programs. The stark reality is that these programs have infiltrated many international courts with but a few independent and auditable investigations carried out on their appropriateness.

While these developments are still evolving, the existence of research and debate should serve as guidelines and cautionary reminders of the necessity for accountability and transparency in this marriage. Technology has breached the global divide and 'ignorance is not an excuse'. There is a responsibility on the judiciary, relating to how technology is disrupting the status quo, not only to welcome the efficiencies promised by predictive systems, but also to take on the role of 'gatekeepers' by ensuring that these systems do not overtake the constitutional values and objectives they aim to protect.

BIBLIOGRAPHY:

Legislation:

- The Constitution of the Republic of South Africa Act 108 of 1996
- Protection of Personal Information (POPI) Act 4 of 2013
- The Criminal Procedure Act 51 of 1977
- Uniform Rules of Court in GN R315 GG19834 of 12 March 1999 (as amended)

International legislation:

- Constitution of the United States of America, 1787- Amendment 14
- General Data Protection Regulation (GDPR) EU Regulation (2016/679) of the European Parliament and of the Council of 27 April 2016

Case law:

- *Association belge des Consommateurs Test-Achats ASBL & Others v Conseil des ministres* C-236/09
- *Levenstein and Others v Estate of the late Sidney Lewis Frankel and Others* [2018] ZACC 16
- *S v Makwanyane and another* 1995 (3) SA 391 (CC)
- *S v Malgas* 2001 (2) SA 1222 (SCA)
- *South African Association of Personal Injury Lawyers v Heath* 2001 (1) SA 883 (CC)
- *State v Loomis* 881 N.W. 2d 749 (Wis 2016)

Books:

- F Du Bois '*Willes Principles of South African Law*' 9th ed (2006)
- Frank Pasquale 'The Black Box Society: the secret algorithms that control money and information' (2015)

Journal Articles:

- B. Alarie, A. Niblett & A. Yoon 'Law in the Future' (2017) 3 *Univ. of Toronto LJ* (Focus Feature)
- Daniella Citron 'Technological due process (2008) 85(6) *Washington University LR* 1253
- Kate Crawford & Jason Schultz 'Big data and due process: toward a framework to redress predictive algorithms, (2014) 55 (1) *Boston Law Review* 93
- Dennis Davis 'Dworkin: A viable theory of adjudication for the South African constitutional community?' 2004 *Acta Juridica* 96
- Guido Noto La Diega 'Against the dehumanisation of decision-making.' (2018) 9 *JIPITEC* 3
- Thomas Flanagan 'Insurance, Human Rights and Equality Rights in Canada: When is discrimination reasonable?' (1985) 18 *Canadian J. of Political Science* 715
- W. Flores, Anthony & Bechtel, Kristin & Lowenkamp, Christopher. 'False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And it's biased against Blacks.' (2016) 80 *Federal probation* 1

- Katherine Freeman 'Algorithmic injustice: how the Wisconsin Supreme Court failed to protect due process rights in State v. Loomis' (2016) 18 *North Carolina Journal of law & Technology* 75
- Willem H. Garvett 'The myth of objectivity: implicit racial bias and the law (Part 2)' (2017) 20 *PER* 1
- Willem Garvett 'Towards an algorithmic model of judicial appointment: the necessity for radical revision of the Judicial Service Commission's interview procedures' (2017) 80 *THRHR* 267
- Pamela Katz 'Expert Robot: using artificial intelligence to assist judges in admitting scientific expert testimony' (2014) 24 (1) *ALB. L.J. SCI. & TECH.* 1
- Joshua Kroll, Joanna Huey Solon Barocas et al 'Accountable Algorithms' (2016) 165 *University of Pennsylvania Law Review* 632
- Matthew T. Nowacheck 'Why robots can't become racist and, why humans can.' (2014) 9 (1) *Phaenex Journal* 57
- K. O'Regan 'Checks and Balances reflections on the development of the doctrine of separation of powers under the South African constitution.' (2005) 8 *PER* 120
- Julius Stone 'Man and machine in the search for justice.' (1964) 16 *Stanford Law Review* 515
- A. Singh and M.Z. Bhero 'Judicial law-making: unlocking the creative powers of judges in terms of section 39(2) of the Constitution.' (2016) 19 *PER* 1
- Julie Sobowale "Beyond Imagination: How artificial intelligence is transforming the legal profession." (2016) *ABA Journal* 47

- Robert J. Spagnoletti 'Using artificial intelligence to aid in the resolution of socioscientific disputes: a theory for the next generation.' 1987 2 *Journal of Law and Technology* 101
- Tene O. & Polonetsky J. 'Taming the Golem: challenges of ethical algorithm decision making' (2017) 19 *North Carolina Journal of Law & Technology*, 125.
- Louise Vincent 'Virginity testing in South Africa: Re-traditioning the postcolony' (2006) 8 *Culture health & Sexuality* 18

Online Articles/ websites:

- Angwin J. Larson J, Mattu S. & Kirchner L. 'Machine Bias' May 2016 available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Owen Bowcott. 'Revealed: divorce software error hits thousands of settlements.' December 2015 available at: https://www.theguardian.com/law/2015/dec/17/revealed-divorce-software-error-to-hit-thousands-of-settlements
- Nicholas Davis 2016 'What is the fourth industrial revolution' World Economic Forum available at: https://www.weforum.org/agenda/2016/01/what-is-the-fourth-industrial-revolution/
- Ron Friedmann 'Automating legal advice: AI and expert systems' available at: https://prismlegal.com/automating-legal-advice-ai-and-expert-systems/
- Gendernet Practice Notes 'Employment, Labour, Social Affairs: Social insurance and equity between women and men.' Available at: www.oecd.dac/gender

- Will Knight 'Biased algorithms are everywhere and no one seems to care' July 2017 available at: https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/
- Dan Kopf. 'This is how computers "predict the future". available at: https://qz.com/1261817/predictive-algorithms-are-not-all-that-complicated/
- Larson J. Mattu S. Kirchner L. & Angwin J. 'How we analysed the COMPAS Recidivism algorithm' May 2016. available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- National Geographic. 'Fulani Initiation Rites' available at: https://video.nationalgeographic.com/video/benin_fulaniinitiation
- National Geographic 'What is a black box?' available at: https://www.nationalgeographic.com.au/engineering/what-is-a-black-box.aspx
- Propublica 'About' available at: https://www.propublica.org/about/
- Redmiles A. & Weller A. (2018) 'Human perceptions of fairness in algorithmic decision making: a case study of criminal risk prediction' available at https://www.researchgate.net/publication/323444031_Human_Perceptions_of_Fairness_in_Algorithmic_Decision_Making_A_Case_Study_of_Criminal_Risk_Prediction
- Klaus Schwab 2016 'The fourth Industrial Revolution: what it means.' World Economic Forum. Available at: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond

- Julie Sobowale ABA Journal *"Beyond Imagination: How artificial intelligence is transforming the legal profession."* Available at: http://www.abajournal.com/magazine/article/how_artificial_intelligence_is_transforming_the_legal_profession
- Technopedia Dictionary 'Algorithm' available at: https://www.techopedia.com/definition/3739/algorithm
- Ed Yong 'A popular algorithm is no better at predicting crimes than random people' 2018 available at: https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

531382:Nokuthula_Olorunju_LLM_Final_2019.pdf

ORIGINALITY REPORT

9%	7%	6%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Oguttu, Annet Wanyana. "Offshore trusts and income tax avoidance: the lures and pitfalls from a South African perspective", International Journal of Private Law, 2012. Publication	1%
2	www.conniefinleyrealestate.com Internet Source	1%
3	Submitted to University of Cape Town Student Paper	1%
4	LexisNexis Publication	<1%
5	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
6	www.laurenkohn.co.za Internet Source	<1%
7	book.douban.com Internet Source	<1%
	Submitted to University of Nottingham	

8	Student Paper	<1%
9	Submitted to University of Ulster Student Paper	<1%
10	Submitted to The University of Law Ltd Student Paper	<1%
11	Submitted to University of Lancaster Student Paper	<1%
12	scholar.sun.ac.za Internet Source	<1%
13	Submitted to Victoria University Student Paper	<1%
14	wiredspace.wits.ac.za Internet Source	<1%
15	media.nesta.org.uk Internet Source	<1%
16	Submitted to Pembroke Student Paper	<1%
17	Submitted to University of Waikato Student Paper	<1%
18	jultika.oulu.fi Internet Source	<1%
19	Submitted to University of Stellenbosch, South	<1%

	Africa Student Paper	
20	Submitted to University of Cambridge Student Paper	<1%
21	www.lifesettlementlawyer.com Internet Source	<1%
22	Submitted to Da Vinci Institute Student Paper	<1%
23	queensu.ca Internet Source	<1%
24	Submitted to Cardiff University Student Paper	<1%
25	Submitted to Laureate Higher Education Group Student Paper	<1%
26	Submitted to University of London External System Student Paper	<1%
27	www.jipitec.eu Internet Source	<1%
28	www.hastingslawjournal.org Internet Source	<1%
29	Submitted to Korea University Student Paper	<1%
30	Submitted to University of Zimbabwe	

—	Student Paper	<1%
31	Submitted to University of Queensland Student Paper	<1%
32	papers.ssrn.com Internet Source	<1%
33	ajol.info Internet Source	<1%
34	repository.up.ac.za Internet Source	<1%
35	Submitted to University of Leeds Student Paper	<1%
36	Submitted to Bournemouth University Student Paper	<1%
37	Submitted to University of Pretoria Student Paper	<1%
38	Submitted to University of Edinburgh Student Paper	<1%
39	www.spicecsm.com Internet Source	<1%
40	Submitted to University of New South Wales Student Paper	<1%
41	Megan T. Stevenson, Christopher Slobogin.	<1%

	"Algorithmic risk assessments and the double-edged sword of youth", Behavioral Sciences & the Law, 2018 Publication	
42	www.vanguardngr.com Internet Source	<1%
43	www.usenix.org Internet Source	<1%
44	humanities-digital-library.sas.ac.uk Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		