

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 uncharted territories. I also extend immense gratitude to my parents for their support.

DECLARATION

I, NOKUTHULA OLOBUNJU

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 Revolution¹ 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. PREDICTIVE 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 Australia²⁴ and Canada²⁵ 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

³³S165(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 Rights⁴⁵- 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