JD (JURIS DOCTOR)

The impact of algorithms in criminal sentencing on due process rights

Anderson, Caoimhe

Award date:
2019

Awarding institution:
Queen's University Belfast

Link to publication

Terms of use
All those accessing thesis content in Queen's University Belfast Research Portal are subject to the following terms and conditions of use

• Copyright is subject to the Copyright, Designs and Patent Act 1988, or as modified by any successor legislation
• Copyright and moral rights for thesis content are retained by the author and/or other copyright owners
• A copy of a thesis may be downloaded for personal non-commercial research/study without the need for permission or charge
• Distribution or reproduction of thesis content in any format is not permitted without the permission of the copyright holder
• When citing this work, full bibliographic details should be supplied, including the author, title, awarding institution and date of thesis

Take down policy
A thesis can be removed from the Research Portal if there has been a breach of copyright, or a similarly robust reason.
If you believe this document breaches copyright, or there is sufficient cause to take down, please contact us, citing details. Email: openaccess@qub.ac.uk

Supplementary materials
Where possible, we endeavour to provide supplementary materials to theses. This may include video, audio and other types of files. We endeavour to capture all content and upload as part of the Pure record for each thesis.
Note, it may not be possible in all instances to convert analogue formats to usable digital formats for some supplementary materials. We exercise best efforts on our behalf and, in such instances, encourage the individual to consult the physical thesis for further information.

Download date: 17. May. 2024
The Impact of Algorithms in Criminal Sentencing on Due Process Rights

Caoimhe Marie Anderson
B.A. Honours, Criminal Justice
Mount Royal University (Canada)

This dissertation is submitted for the degree of the Juris Doctor (Law)
Queen’s University Belfast (Northern Ireland)
September 2019
DISCLAIMER

This thesis submission complies with Queen’s University Belfast’s Principles for Professional Doctorates, The Study Regulations for Research Degree Programmes, the Code of Conduct and Integrity in Research, the Intellectual Property Policy, and policies and procedures for research ethics, including those applying in the School of Law. This project has been approved by the School of Law Ethics Committee, and the primary supervisor as a ‘low risk' project. In accordance with the School of Law guidelines, this thesis does not exceed 50,000 words. (The Portfolio does not exceed 10,000 words and the dissertation does not exceed 40,000 words).

All errors are my own. This law is current as of September 13, 2019.
Acknowledgments

I would like to acknowledge the staff at the School of Law for all of their support and guidance throughout the Juris Doctor program. In particular, I must thank my supervisor, Daithí MacSithigh, for his continuous support and advice. His thought-provoking comments helped me better this project and my writing throughout. Thank you also to my secondary supervisor, Cheryl Lawther – her feedback has helped polish this piece significantly. A special ‘thank you’ to Heather Kinning for convincing me to come back to my homeland to study law – it genuinely has been some of the best years of my life thus far. I am thankful for all of the wonderful people that I have met during my time in Northern Ireland – it wouldn’t have been the same without each and every one of you. I must acknowledge my former professor, Beverley McEwan for guiding me towards obtaining a university education and pushing me out of my comfort zone. Thank you to my papa who encouraged me to keep working hard during our visits every Sunday. Finally, a massive thank you to my parents and brother who have supported me through the highs and lows of my university career – I don’t know what I would do without you.
Abstract

There is a growing trend towards the adoption of algorithms and automated data processing systems in areas of life which were previously entirely in the remit of human beings. The criminal justice system and in particular, criminal sentencing, is no exception. The use of algorithmic tools in criminal sentencing creates various implications for due process rights due to their opacity and the use of particular input variables.

Currently, the majority of academic literature focuses on the use of algorithms in the criminal justice system, but largely ignores their use in criminal sentencing. This project utilizes a socio-legal approach and applies a human rights lens to analyse the impact of algorithms in criminal sentencing on due process rights. This project addresses four sub-questions: How are algorithms and automated data processing systems utilized in the context of criminal sentencing? How does the use of algorithms in criminal sentencing impact due process rights? What safeguards are available to those subject to automated data processing at criminal sentencing and how effective are they? What, if any, amendments could be made to enhance the protection of due process rights from algorithmic technology?

Currently, courts have demonstrated little restraint regarding the use of algorithmic tools in criminal sentencing, despite the lack of understanding of algorithms and the fact that these tools can result in biased and discriminatory outputs. The use of algorithms in criminal sentencing can have serious implications on due process rights including the right to review and verify sentencing, the right to an individualised sentence, and the right to equality before and under the law. The current safeguards, in the form of data protection legislation, do not effectively protect defendants due process rights who are subject to algorithmic tools at criminal sentencing. This project recommends that various safeguards must be put in place should other countries wish to follow the trend of adopting algorithms at criminal sentencing. These safeguards include the adoption of a human-rights framework, expert testimony, regular audits, enhanced technological transparency through open source codes, and mandatory judicial training on assessing algorithmic evidence.
# TABLE OF CONTENTS

**Introduction** p. 6

**This Project** p. 11

- Structure p. 12
- Methodology p. 15
- Methods p. 17
- Challenges and Limitations p. 17

**Chapter 1: The Use of Algorithms in the Context of Criminal Sentencing** p. 19

- Background p. 19
- Theories of Sentencing p. 22
- The Role of the Automated Data Processing System in Sentencing p. 24
- Studies on Algorithmic Tools in Sentencing p. 27
  - *State of Wisconsin v. Eric L. Loomis* p. 33
- Non-American Case Law p. 36
- Statistical Accuracy p. 43
- Summary p. 43

**Chapter 2: The Implications on Due Process Rights** p. 46

- What are Due Process Rights? p. 47
- The Human Rights Issues Implicated By Risk Assessment Algorithms p. 51
- The Right to Review and Verify Sentencing Information p. 51
- The Right to an Individualised Criminal Sentence p. 55
- Algorithmic Input Variables and the Implication on Due Process p. 57
- Race as an Input Variable p. 57
- Gender as an Input Variable p. 61
- Socioeconomic Status as an Input Variable p. 66
- Education & Employment as Input Variables p. 68
- Counterarguments in Defence of Algorithms in Criminal Sentencing p. 69
  - The “Advisory Tool” Argument p. 69
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Exclusion of Input Variables at Cost of Statistical Accuracy</td>
<td>70</td>
</tr>
<tr>
<td>Conclusion</td>
<td>71</td>
</tr>
<tr>
<td>Chapter 3: Strengths and Weaknesses of the Current Safeguards</td>
<td>74</td>
</tr>
<tr>
<td>The General Data Protection Regulation (GDPR)</td>
<td>75</td>
</tr>
<tr>
<td>The right not to be subject to ‘solely’ automated decisions</td>
<td>76</td>
</tr>
<tr>
<td>Notification Duties &amp; Envisaged Consequences</td>
<td>79</td>
</tr>
<tr>
<td>Accurate, Adequate, Relevant and Necessary</td>
<td>80</td>
</tr>
<tr>
<td>Does the GDPR address discrimination?</td>
<td>81</td>
</tr>
<tr>
<td>The Law Enforcement Directive</td>
<td>83</td>
</tr>
<tr>
<td>The Data Protection Act 2018</td>
<td>86</td>
</tr>
<tr>
<td>Are the GDPR, LED and DPA adequate safeguards?</td>
<td>86</td>
</tr>
<tr>
<td>United States Data Protection</td>
<td>88</td>
</tr>
<tr>
<td>Are the current safeguards effective?</td>
<td>90</td>
</tr>
<tr>
<td>Conclusion</td>
<td>91</td>
</tr>
<tr>
<td>Chapter 4: Recommendations &amp; Safeguards</td>
<td>92</td>
</tr>
<tr>
<td>Removal or Prohibition of Algorithms from Criminal Sentencing</td>
<td>93</td>
</tr>
<tr>
<td>Implementation of Mandatory Judicial Training on Assessing Algorithms</td>
<td>96</td>
</tr>
<tr>
<td>Expert Testimony on Algorithms</td>
<td>99</td>
</tr>
<tr>
<td>Technological Transparency: Open Source Code &amp; Rule Explanation</td>
<td>101</td>
</tr>
<tr>
<td>Regular Audits</td>
<td>104</td>
</tr>
<tr>
<td>Designing Algorithms with Due Process &amp; Criminal Sentencing in mind</td>
<td>106</td>
</tr>
<tr>
<td>ALGO-CARE</td>
<td>107</td>
</tr>
<tr>
<td>Practicality &amp; Efficacy of these Safeguards</td>
<td>110</td>
</tr>
<tr>
<td>Conclusion</td>
<td>112</td>
</tr>
<tr>
<td>Conclusion</td>
<td>114</td>
</tr>
<tr>
<td>The Research Outlook</td>
<td>117</td>
</tr>
<tr>
<td>Bibliography</td>
<td>118</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>135</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>152</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>168</td>
</tr>
</tbody>
</table>
The Impact of Algorithms in Criminal Sentencing on Due Process Rights

Providing equal justice for poor and rich, weak and powerful alike is an age-old problem. People have never ceased to hope and strive to move closer to that goal… In this tradition, our own constitutional guarantees of due process and equal protection both call for procedures in criminal trials, which allow no individual discriminations… The central aim of our entire judicial system is that all people charged with crime must, so far as the law is concerned, stand on the bar of equality before the bar of justice in every court.

--- Justice Hugo Black

Over the past forty years, computer technology has dramatically reformed legal practice. It is now used to draft and communicate legal advice, correspondence and submissions, as well as to facilitate primary and secondary research, and thereby enhance the efficiency and accuracy of legal advice and judgments. Currently, however, computers have had relatively little impact on judicial decision-making. This may be unsurprising given that judicial sentencing demands high-level, complex reasoning and thought, and is at present greatly informed by human observation, impression and intuition. Lately, however, various news outlets have reported on actors within the court system being replaced by ‘robots’ and computer software. Academic research has also been conducted on the proposition of lawyers and judges being replaced by ‘robots’ or

---

2 Mirko Bagaric & Gabrielle Wolf, ‘Sentencing by computer: enhancing sentencing transparency and predictability, and (possibly) bridging the gap between sentencing knowledge and practice,’ 2017 George Mason Law Review 25(4) 1-50 p. 21
3 Ibid.
4 Ibid.
Although there is no real concern of judges being replaced by computer software within the next week, computers are gradually beginning to play a meaningful role in complex areas of law, including judicial sentencing and decision-making. This is principally due to the development of increasingly sophisticated algorithms.\(^6\)

The term ‘algorithm’ comes from computer science, rather than law, and refers to an automatic rule that uses numerical inputs to produce a result. In practice, it refers to the automation of a statistical method.\(^8\) The Pew Research Center notes that algorithms are not new, ‘they are simply instructions of solving a problem or completing a task.’\(^9\) The increasing use of algorithms, or risk assessment tools, stems from the fact that, at present, ‘massive amounts of data are being created, captured and analyzed by businesses and government’s.’\(^10\) Algorithms do not only enable Internet users to search for and access information, but they are also increasingly being used in decision-making processes that were previously entirely in the remit of human beings. Algorithms may be used to assist human beings with decisions or to take them immediately through automated means. ‘In fact, boundaries between automated and human decision-making are often blurred, resulting in the notion of ‘quasi-or semi-automated decision-making.’\(^11\) Machine learning has advanced so much that it has now progressed to the point where accurate predictive models can be developed using word sequences to ascertain factors that influence the outcomes of judicial decisions. This can help lawyers and members of

\(^6\) See for example: John Morison & Adam Harkens, ‘Re-engineering justice? Robot Judges, Computerized Courts and (semi) automated legal decision-making,’ 2019 Forthcoming Legal Studies
\(^7\) Bagaric & Wolf (n 2) p. 21
\(^10\) Ibid.
\(^11\) Committee of experts on internet intermediaries (MSI-NET), ‘Study on the human rights dimension of automated data processing techniques (in particular algorithms) and possible regulatory implications,’ 2017 Council of Europe accessed: https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a p. 3
the judiciary to identify efficiently relevant cases and patterns that drive legal decision-making.\textsuperscript{12}

An increased reliance on predictive algorithms, more commonly known as risk assessment tools, takes place within a wider context of mass incarceration and racial discrimination, particularly within the United States criminal justice system.\textsuperscript{13} The United States, which represents five per cent of the world population, has 25 per cent of the world’s prisoners: 2.2 million people (including pre-trial detainees) are currently incarcerated, and one in twelve black men between 25 and 56 years old are currently imprisoned.\textsuperscript{14} There is a similar trend in Canada, the United Kingdom and many other jurisdictions.\textsuperscript{15} ‘Part of what is driving the introduction of algorithms in courts is the hope that these tools will have a positive impact on the overall fairness of the system.’\textsuperscript{16} Bagaric and Wolf argue that computerized sentencing is preferable to judicial sentencing because it will eliminate judge’s subconscious bias on sentencing decisions and achieve decisions that uphold the rule of law values of consistency, transparency and predictability.\textsuperscript{17} In contrast, others have argued that although judicial reasoning may be subconscious and seemingly opaque, algorithms introduce a form of opacity that is not readily intelligible to humans.\textsuperscript{18}

Reservations against algorithms and automated data processing techniques in criminal sentencing usually point to their opacity and unpredictability. However, there is an increasing realization that specific human rights are particularly affected.\textsuperscript{19} The use of algorithms and automated data processing techniques raises concerns for fair trial standards contained within various human rights instruments including the ECHR and

\textsuperscript{12} Bagaric & Wolf (n 2) p. 22
\textsuperscript{13} Christin, Rosenblat & Boyd (n 8) p. 13
\textsuperscript{14} Ibid.
\textsuperscript{16} Christin, Rosenblat & Boyd (n 8) p. 2
\textsuperscript{17} Bagaric & Wolf (n 2)
\textsuperscript{18} Jenna Burrell, ‘How the machine ‘thinks’: Understanding opacity in machine learning algorithms,’ 2016 \textit{Big Data & Society} 10 1-12
\textsuperscript{19} Committee of experts on internet intermediaries (n 11) p. 10
national constitutions. Examples of the human rights that may be impacted include Article 6 of the ECHR, which denotes the right to be informed of the cause and nature of the accusation, and the right to a fair hearing and the presumption of innocence. Concerns may also arise with respect to Article 5 of the ECHR, which protects against arbitrary deprivation of liberty, and Article 7, which outlines that there is no punishment without law. A further example of a human rights instrument that could be impacted by the use of algorithms in criminal sentencing would be the United States Constitution, particularly the Fifth Amendment. Despite the reservations against the use of algorithms and automated data processing systems in criminal sentencing, some states have declared their use in sentencing to be constitutional.

This research project will explore the use of automated data processing systems, such as algorithms, in the context of criminal sentencing and analyze the impact of the use of these systems on due process rights drawing on examples from various countries and particularly the United States given the popularity and growing acceptance of their use in U.S. jurisdictions. Furthermore, legal scholars have argued for years that automated processing requires greater transparency, but it is far from clear what form such transparency should take. The most obvious approach that has been suggested is to disclose the source code of a system. However, these codes are often illegible to non-experts and in many cases, and the code only exposes the machine learning method and not the data-driven decision rule. Moreover, transparency may be undesirable in particular situations, thus making partial opacity preferable to prevent individuals from gaming the system. This project will also attempt to uncover the form which ‘transparency’ should take concerning automated data processing systems used in criminal sentencing in order to best protect individual due process rights.

---

20 Committee of experts on internet intermediaries (n 11)
23 Ibid.
Finally, this project will analyze various case law and legislation designed to protect individuals from algorithmic harms, in particular, the latest legislation adopted by the European Union - the General Data Protection Regulation. This project will address the efficacy of this legislation in the context of protection of due process rights, and to attempt to answer, what, if any, amendments could be made to enhance the protection of due process rights from algorithmic technology. In short, this project seeks to make several contributions to the existing critical literature on algorithmic technology used in criminal sentencing. Specifically, it will provide a human rights-based critique; an exploration of the socio-legal consequences of algorithmic technology; and a critical examination of the underlying empirical research and instruments relevant to protection from algorithmic harms. Lastly, this project will provide recommendations and considerations for the future protection of due process rights from the use of automated data processing systems in criminal sentencing.

This Project

This project seeks to address four primary questions:

(1) How are algorithms and automated data processing systems utilized in the context of criminal sentencing?
(2) How does the use of algorithms in criminal sentencing impact due process rights?
(3) What safeguards are available to those subject to automated data processing at criminal sentencing and how effective are they?
(4) What, if any, amendments could be made to enhance the protection of due process rights from algorithmic technology?

At present, the literature focuses heavily on the implications of using algorithms in the areas of the criminal justice system other than criminal sentencing, such as policing
and rehabilitation. Any academic that has narrowed their focus to the due process implications of algorithms used in criminal sentencing has provided analysis exclusively on case law, constitutional rights, and jurisprudence stemming from the United States. This is unsurprising given the popularity of the use of algorithms in criminal sentencing in the United States; however, there have been 86 court closures in England and Wales, with a further 15 identified for further action due to ‘online court’ processes. Other European countries are also adopting algorithmic tools and they are often welcomed as a way to cope with diminished resources and the ongoing quest to increase public safety. Moreover, there is a lack of academic literature evaluating the currently available safeguards for algorithmic harms in the context of due process rights, even in United States literature. These gaps in the current literature, combined with the growing trend of algorithm usage in courts, provided justification for both the subject matter and scope of this research.

Structure

The first chapter of this dissertation will provide an answer for the first question: (1) how are algorithms and automated data processing systems utilized in the context of criminal sentencing? This chapter comprises of a review of the available literature to provide the reader with both of a context of how these tools are used and identify the arguments and counterarguments set forth by academics on the use of algorithms in criminal sentencing and the criminal justice system more generally. Chapter 1 begins by presenting studies conducted by academics researching the use of algorithms in the criminal justice system more generally, for example, risk assessment tools in the context of policing and bail decisions. These studies often focused on the reliability and accuracy of these algorithmic tools and points to whether they produce seemingly biased results.

The beginning of this chapter also provides the reader with information on what variables are used as inputs into algorithmic programs and how these variables can affect the output. This chapter then moves to discuss theories of criminal sentencing and the theoretical justification or lack thereof for the use of algorithms in criminal sentencing. Chapter 1 then discusses the literature regarding the use of algorithms in criminal sentencing specifically and some examples of case law wherein the use of these algorithmic tools in sentencing were challenged. This chapter concludes with a discussion on the construction of these algorithmic tools and the problems that arise from the opacity of the tools themselves and the actors that contribute to their construction. This chapter sets the scene by providing the reader with an understanding of the various problems that can arise out of the use of algorithms in criminal sentencing.

Chapter 2 seeks to answer the second question: (2) how does the use of algorithms in criminal sentencing impact due process rights? Chapter 2 begins by providing a definition of ‘due process rights’ and a review of the literature and the articles or sections of human rights instruments regarding due process rights. Although this project is focused on algorithmic technology and criminal sentencing, the work is situated in general due process and human rights literature. In other words, this project is not evaluating the impact of algorithms in criminal sentencing based on a newfound definition of due process rights, but rather, due process rights, as they are understood in the traditional context of human rights law. Chapter 2 then moves onto a discussion of the challenges that arise out of algorithms being used in criminal sentencing. The first identified due process challenge is the right to verify and review sentencing information. The second challenge discussed is the right to an individualized sentence and how algorithms may interfere with this right. This chapter then discusses the due process challenges associated with the use of particular input variables and the effects that they can have on a criminal defendant’s sentence. The variables are broken down into particular subsections such as race, gender, socioeconomic status, education and employment. Finally, this chapter analyses the arguments and counterarguments for the inclusion of these variables as inputs in algorithmic tools and concludes that the use of these tools can present various challenges to due process rights that must be addressed.
Chapter 3 answers the third question: (3) What safeguards are available to those subject to automated data processing at criminal sentencing and how effective are they? The General Data Protection Regulation (GDPR) 2016/679 is the first safeguard discussed, as there was great anticipation among the academic community that this would provide those subject to algorithms and automated data processing systems with various rights and protections. The second safeguard discussed is another piece of EU legislation, parallel to the GDPR, known as Directive (EU) (2016/680) or the ‘Law Enforcement Directive’ (LED). The LED deals with the processing of personal data by data controllers for law enforcement purposes, which fall outside the scope of the GDPR. One example of the LED’s transposition into domestic law will also be analyzed (The United Kingdom’s Data Protection Act 2018). Each of these statutory instruments is analyzed for efficacy in the context of protecting due process rights from algorithms and automated processing. This analysis is conducted only on the sections and articles pertinent to processing in the criminal sentencing context and the protection of due process rights as an analysis of these legislative instruments in their entirety would be beyond the scope of this project. Chapter 3 also provides an analysis and evaluation of the limited safeguards available to those in the United States given the popularity of the use of algorithms in courtrooms throughout that jurisdiction. Examples include California’s recently implemented data privacy protection law known as the California Consumer Privacy Act (CCPA). Chapter 3 concludes that the current safeguards do not provide adequate protection of due process rights.

The final chapter of this project will provide recommendations on safeguards that could provide better protection for defendant’s due process rights when subject to algorithms or automated data processing systems at a criminal trial. This chapter also

---

28 See Chapter 3
provides an argument for the removal of algorithmic tools from criminal sentencing altogether, however, as it is expected that the trend of adopting these tools will increase, this chapter lays out recommendations that both policymakers, courts, and algorithm developers may wish to consider before proceeding. The recommended safeguards include statutory amendments and additions, procedural practices and preventative measures designed to address the problems with algorithms in sentencing before they have the opportunity to arise.

Note that there is an Appendix within this dissertation project that includes two further documents on this topic. One of these documents is a case note which includes an in-depth analysis of the United States case *State of Wisconsin v. Loomis* [2016] 881 N.W.2d 749 (Wis. 2016). The primary justification for choosing this case was that it is currently one of the only cases wherein a Court has been challenged for utilizing an algorithmic tool (COMPAS) in sentencing on the grounds that it violated a defendant’s due process rights.29 This case is also discussed throughout the dissertation, including in Chapter 3 wherein safeguards such as the GDPR are applied to the case in order to understand the hypothetical outcomes. The other document is a legislative report that was written as a preparatory piece for this project. The legislative report is a detailed analysis of particular sections of the General Data Protection Regulation (EU) 2016/679 (GDPR) pertaining to rights that could potentially be invoked by those subject to data processing at trial. The GDPR and its related Directives are also evaluated as a safeguard in Chapter 3. The justification for selecting the GDPR was due to its wide territorial scope and the creation and enhancement of rights that have previously not been available to data subjects.30

**Methodology**

This project utilizes an interdisciplinary approach to studying algorithms and their impact, within legal studies. The primary research method employed was socio-legal

---

29 See Case Note in Appendix
30 See Chapter 3 and Legislative Report
research with the application of a human rights lens. Socio-legal methods employ sociology and social science solely for data collection purposes; thus, socio-legal researchers use social theory as the basis of analysis for law and legal studies not related to concerns in sociology or other social sciences.\(^{31}\) This interdisciplinary approach is appropriate because both sides can inform and develop new knowledge about algorithms and the law’s designed to address them, and their impact on due process and society that could not have been grasped sufficiently alone.\(^{32}\) A strictly doctrinal approach would not have worked for this piece given both the broad scope of the current legislation that is applied to this topic and also the current lack of case law and legislation available regarding this topic.

This project also applied a ‘human rights lens’ for conceptualizing and implementing a research and evaluation project.\(^{33}\) Although the project primarily focused on what are defined as ‘due process rights,’\(^{34}\) human rights research methodology flows throughout the entirety of the project, including in the recommendations for future safeguards in Chapter 4. To conduct research following a human rights approach, there are strategies that must be followed. First, there must be an understanding and application of the values and principles of the human rights framework.\(^{35}\) These principles include the principles of universality, non-discrimination, indivisibility and interdependence, participation, accountability and transparency.\(^{36}\) Although all human rights principles are important, the principles of non-discrimination and transparency were particularly relevant to this project given that the main problems identified through the literature with algorithms were their inherent opacity and potentially discriminatory outputs.

The human rights ‘principle of accountability suggests that governments are responsible for creating a mechanism of accountability for the enforcement of equal rights, which includes monitoring and evaluating the implementation of laws and policies to protect

---

\(^{31}\) Reza Banakar & Max Travers, *Theory and Method in Socio-Legal Research* (Hart Publishing 2005) xi-xii

\(^{32}\) Ibid.


\(^{34}\) See Chapter 2

\(^{35}\) Maschi (n 33)

\(^{36}\) Maschi (n 33)
This principle was the underlying force in the conception of the recommended future safeguards in Chapter 4 of this project and throughout the analysis of currently available safeguards in Chapter 3.

Methods

The formal methods that were used to create this project were qualitative, meaning they did not depend on statistical quantification but attempted to capture and categorize social phenomena and its meanings. This project was ‘desk-based’ and involved an extensive review of the literature. The primary literature reviewed included legislation and case law, while secondary data collected included academic journals, government reports, newspaper articles and regulatory body reports such as the Information Commissioners Office, etc.). Academic journals that were analyzed generally, but not exclusively, focused on ‘law’ and ‘technology,’ while the case law that was analyzed primarily dealt with cases involving a complaint of human rights infringements. The areas of discipline in which the majority of the literature was gathered from included computer science and technology, jurisdictional law reviews, privacy and security and human rights law. The data gathered from computer science and technology journals provided a greater insight into the structure and composition of algorithms and assisted in developing the recommendations in Chapter 4. The data gathered from the law and human rights journals assisted in analyzing the due process implications of the use of algorithms. Overall, the review of primary legal data and secondary data was necessary to enhance the understanding of how algorithms are used in criminal sentencing and how they can have an impact on individual due process rights.

Challenges and Limitations

Various challenges occurred while attempting to address this topic. First, the majority of the literature that is currently available on the topic of algorithms in criminal

---

37 Maschi (n 33)
sentencing stems from the United States. There was very limited empirical data on the use of algorithms in criminal sentencing outside of the United States as many countries have yet to implement them in sentencing specifically. Further, the majority of empirical studies conducted on algorithmic tools focus on their use in a policing context, rather than sentencing. Moreover, the current safeguards were created by legislature’s within the European Union and despite the GDPR’s wide scope, its provisions are not directly applicable to citizens from outside of the EU. Thus, the application of these safeguards to individuals subject to algorithms at trial was hypothetical throughout this project and based on the presupposition that algorithms may become a standardized tool in criminal sentencing.

Another limitation is the lack of empirical data on the implications of algorithms on due process rights more generally. When algorithms and automated data processing are the topics of the literature, it is often their implication on privacy rights that is the subject of research. Once again, any literature that provided any form of due process analysis was based on the Constitution of the United States, rather than international human rights instruments. This meant that much of the analysis on the impact of algorithms on due process rights based on international instruments and European instruments was contributory and not based on citing other academic’s works. Hopefully, this project will provide useful groundwork for future research should algorithms become a standardized sentencing tool internationally.
Chapter 1: The Use of Algorithms in the Context of Criminal Sentencing

This chapter will provide an overview of the current literature on the use of algorithms in both the criminal justice system and in sentencing. An examination of case law, academic articles, including empirical studies on algorithmic tools, that are used by police, probation officers and judges in various jurisdictions will be undertaken so as to glean a better understanding of their impact and justifications for their use in criminal sentencing. Moreover, this chapter will also discuss the theories of sentencing and the arguments and counterarguments for the use of algorithmic tools in sentencing. Finally, this chapter will provide oversight of the structure and construction of the algorithmic tools used in sentencing and draw attention to problems that can arise due to their technological constitution.

Background

There are currently more than sixty predictive tools drawing on large amounts of data in the US criminal justice system. The algorithmic tools will typically provide an estimate of an offender’s risk of recidivism or failure to appear in court when on bail, generally expressed in a range of ‘low’ to ‘high’ risk. These estimates are based on a small number of variables about defendants, either connected to their criminal histories (violent offences, previous offences, failure to appear in court, etc.) or socio-demographic characteristics (e.g. sex, history of drug use, employment status, education, etc.). These predictive tools are explicitly designed to ‘structure’ the criminal decision-making process and curtail judicial discretion by providing a clear set of guidelines, scores, and recommendations to judges, prosecutors, and probation officers in charge of making decisions about cases. The United States is not the only country using algorithms to assist with decision-making within the criminal justice system. Kent Constabulary have been using a commercial ‘PredPol’ algorithm since 2013; a ‘predictive policing tool’ to

---

40 Ibid.
41 Ibid.
identify ‘where offences are likely to take place’ using data on past crime patterns.\textsuperscript{42} RUSI highlighted that a similar algorithm developed in-house by the Greater Manchester Police in 2012 had been shown to ‘be effective in reducing burglary.’\textsuperscript{43} The Durham Constabulary is also using an algorithm known as ‘HART’ to assist in assessing whether a suspect could be eligible for a deferred prosecution.

Urwin conducted a study on the algorithmic forecasting of offender dangerousness for police custody officers based on the Harm Risk Assessment Tool (HART) used by Durham Constabulary officers on offenders at the entry point to the criminal justice system. HART forecasts over two years whether a suspect is likely to commit a serious offence over the next two years (high-risk), any offence (moderate risk), or no offences (low-risk). The study found that the error rate of a high-risk false-negative result was 2%. Therefore, Durham Constabulary can be 98% sure that the worst-case scenario will not happen, should custody officers follow the decision to support algorithmic tool forecasting.\textsuperscript{44} The HART model became more conservative in its forecast of high-harm in order to minimize the worst-case scenario of serious harm in our communities.\textsuperscript{45} This study found that police officers were more likely to forecast a “moderate” when deciding on the dangerousness of suspects than the algorithm. Urwin argues that this can result in an out of court disposal for a potentially dangerous suspect, or sentencing an offender to prison that is unlikely over the next two years to be rearrested.\textsuperscript{46} The researchers state that the conclusions drawn by the HART model are probable but not conclusive, recognized by the advisory nature of the algorithm.\textsuperscript{47}


\textsuperscript{43} Alexander Babuta, ‘Big Data and Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities,’ 2017 \textit{Royal United Services Institute for Defence and Security Studies} 38(2) p. 20


\textsuperscript{45} Ibid

\textsuperscript{46} Ibid. p.86

\textsuperscript{47} Ibid. p. 22
Lum and Isaac applied a policing algorithm, developed by PredPol, one of the largest vendors of predictive policing systems in the USA, to police data collected from a synthetic population of the residents of the city of Oakland. Its founders described the algorithm as a parsimonious race-neutral system that uses only three variables to make predictions: location of the crime, time of the crime, and past type of crime. It was found that rather than correcting or reversing the apparent biases in police data, the algorithm reinforced these biases. Using the PredPol algorithm in Oakland, black people would be targeted by predictive policing at roughly twice the rate of Caucasian people. Individuals classified as a race other than black or white, would receive targeted policing at a rate 1.5 times that of Caucasian people. The researchers held that although predictive policing is simply reproducing the same biases that the police have historically held, filtering this decision-making process through sophisticated computerized software provides unwarranted legitimacy to biased policing strategies, and ultimately affects the right to due process.48 In contrast, Skeem & Lowenkamp found that the use of the “Post Conviction Risk Assessment” (PCRA) tool showed very little predictive bias by race when accurately predicting rearrests of black and white offenders. Overall, this study deemed the use of the PCRA tool as ‘useful for assessing the risk of future crime.’49

Various other studies have been conducted on ‘risk assessment’ algorithmic tools internationally and often produce inconsistent results. In general, research suggests that the most commonly used risk assessment instruments and algorithms can, with a moderate level of accuracy, predict who is at risk for violent recidivism.50 The research also suggests that there are no distinctions in regards to predictive validity of the most commonly used risk assessment tools and algorithms. Although there has been considerable focus on using automated data processing systems and algorithms in rehabilitation and pre-trial decision-making, they have recently drawn attention for their

50 Nathan James, ‘Risk and Needs Assessment in the Federal Prison System,’ 2018 Congressional Research Service supra note 41 at 3
use in sentencing\textsuperscript{51} – the primary focus of this paper. Lawmakers and judges have been relatively quick to embrace the use of algorithms in the sentencing context, particularly within the United States. The ‘successful’ use of these tools in other parts of the justice system may explain this. Sentencing, however, involves a much broader range of considerations than parole and pre-trial risk assessments. A sentencing decision involves deciding how to punish someone, and then, if a judge chooses incarceration, the length of the sentence. Determining the severity and length of punishment often draws upon a number of different theories of punishment, including individual retribution, rehabilitation, deterrence, and incapacitation.\textsuperscript{52}

**Theories of Sentencing**

Theoretical justifications for sentencing differ significantly making generalizations on an international scale impossible. Nonetheless, most western scholars of sentencing distinguish between two broad and opposing approaches to the allocation of criminal punishment. The first of these approaches is usually termed ‘retributive’ or ‘deontological.’ The adherents of this approach believe that an offender’s blameworthiness or culpability for a crime committed in the past should be the only consideration in determining their punishment.\textsuperscript{53} The alternative approach is the ‘consequentialist’ or ‘utilitarian’ approach wherein the adherents of this approach take the position that the effect of punishment on preventing crime by an offender or by others should be the only concern in setting punishment.\textsuperscript{54} ‘Many scholars endorse some form of a hybrid approach to sentencing that includes elements of both the retributive/deontological and the consequentialist/utilitarian theory.’\textsuperscript{55} This hybrid approach is often called ‘limiting retributivism.’\textsuperscript{56}

\textsuperscript{51} Kehl, Guo, & Kessler (n 24) p. 10  
\textsuperscript{52} Ibid. p. 13  
\textsuperscript{54} Ibid.  
\textsuperscript{55} Ibid. p. 492  
\textsuperscript{56} Norval Morris, ‘The Future of Imprisonment,’ 1974 University Chicago Press
Monahan & Skeem are some of the limited numbers of researchers that have explored the area of risk assessment tools in regards to theories of criminal sentencing, and they argue that risk assessment tools are relevant to utilitarian (crime control), but not retributive (just deserts), sentencing concerns. In contrast, Monahan & Skeem’s argument regarding algorithms and risk assessment tools being aligned with utilitarian (crime control) theories of crime is justified as the incapacitation of high-risk offenders and risk education (the rehabilitation of offenders) are of central importance in utilitarian theories. In contrast, they argue that retributive theories are based on perceptions of blameworthiness. Blameworthiness consists of two components: (1) the seriousness of the harm caused to by the crime that the offender has been convicted for, and (2) the mens rea of said crime - i.e. the intent, motive and mental capacity at the time that the crime was committed. Thus, the assessment of future crime through the use of risk assessment tools such as algorithms plays no role in sentencing decisions based solely on backwards-looking perceptions of blameworthiness.

Monahan & Skeem argue that the retributive task of assigning blame for past crime and the utilitarian task of assessing risk for future crime is ‘orthogonal,’ and that it is difficult to integrate these concerns to determine an offender’s sentence when a given factor bears on only one concern such as gender, but not blameworthiness. Or when a risk assessment bears concern in both directions but also in opposite directions - for example combat induced trauma increases risk but can mitigate the perceptions of blameworthiness. Monahan & Skeem also argue that clear conceptualizations and precise terminology are needed to advance the use of risk assessment tools in criminal sentencing.

---

57 Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53)
58 Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53)
59 Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53) p. 492
60 Ibid.
61 Ibid.
62 Ibid.
63 Ibid.
Does the algorithm or risk assessment tool fit into the idealized hybrid theory of ‘limited retributivism?’ Slobogin argues that any risk assessment tool used under this hybrid theory of sentencing means that any very high estimated risk of future crime output does not justify a sentence that exceeds the upper boundary of severity perceived as morally proportionate to the crime of which the offender has been convicted. In other words, a risk assessment tool should not be used to sentence offenders for more time than they morally deserve.64 ‘Ideally, retributive concerns set a permissible range for the sentence, and a risk assessment is only used to select a particular sentence within that range. Risk assessment should not be used to sentence offenders to more time than they morally deserve.’65 This project will demonstrate that the literature on the use of risk assessment tools in sentencing has, however, indicated that there may be situations in which certain individuals could be receiving longer sentences than they morally deserve based on particular attributes they have, or based on the construction of the automated data processing system itself.

The Role of the Automated Data Processing System in Sentencing

In 1963, Reed C. Lawlor speculated that computers would one day become able to analyze and predict the outcomes of judicial decisions.66 Lawlor maintained that, in order to achieve a reliable prediction of judicial activity, there must be a scientific understanding of the ways that the facts in cases and the law impacts on the relevant decision-maker.67 Over fifty years later, the first systematic study to attempt to predict the outcomes of cases decided by the judges of the European Court of Human Rights (ECtHR) focusing solely on textual context was able to predict outcomes with 79 per cent accuracy.68 Previous work on predicting judicial decisions has largely focused on the

64 Christopher Slobogin, ‘Prevention as the primary goal of sentencing: the modern case for indeterminate dispositions in criminal cases,’ 2011 San Diego Law Review 48(2011) 1127-1172
65 Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53) p. 508
67 Ibid
analysis and predictions of judges’ votes given non-textual information, such as the nature and gravity of the crime or the preferred policy position of each judge. The empirical analysis of the ECtHR judicial decisions indicates that the formal facts of a case are the most important factor in accurately predicting judicial decisions. ‘This is consistent with the theory of legal realism suggesting that judicial decision-making is significantly affected by the stimulus of facts.’ The researchers of this study noted that building a text-based predictive system of judicial decisions could offer judges and lawyers with a useful assistance tool as it could be used to rapidly identify cases and extract patterns that correlate with certain outcomes and in the context of the ECtHR, it could be used to develop prior indicators for diagnosing potential violations of specific Articles in lodged applications, ultimately providing a system of prioritization where violations seem particularly likely. Although automated data processing systems are not at a stage to replace judges yet, this study on the ECtHR is a demonstration of how much these systems are progressing and how the boundaries between human and automated decision-making can blur.

Advocates for statistical approaches and the use of algorithms in criminal justice highlight the benefits associated with risk assessment tools in sentencing. Proponents to the use of automated processing systems in the context of criminal sentencing argue that these tools rationalize the decision-making process by summarizing all of the information in a much more efficient way than the human brain could: actuarial assessments would do a better job than individualized judgment at predicting risk and therefore would help strike a better balance between reducing incarceration and promoting public safety. Advocates from the United States also hope that these algorithmic tools will help tackle the history of racism and discrimination embedded in the U.S. criminal justice system by reducing disparities in the interpretation of the information available to judges.

---

70 Alestras et al. (n 68) p. 1
71 Aletras et al. (n 68)
72 Christin, Rosenblat & Boyd (n 8) p. 1
prosecutors and other court staff during the decision-making process.\textsuperscript{73} Data-driven initiatives are frequently said to minimize incarceration rates and the length of imprisonment for low-risk offenders, which results in lower budgetary costs and reduced social harm.\textsuperscript{74} Furthermore, using predictive algorithms may also save time for overworked prosecutors, judges, clerks, and other court staff, however, there is very little research about whether predictive algorithms achieve the goals that those advocates strive for.\textsuperscript{75}

Monahan & Skeem note that there are three primary roles of risk assessment tools in criminal sentencing. The first role is to inform decisions regarding the imprisonment of higher-risk offenders. Risk assessment can provide an empirical estimate of whether an offender has a sufficiently high likelihood of again committing a crime to justify incapacitation within the range of severity set by moral concerns about the criminal act of which the offender has been convicted.\textsuperscript{76} According to Monahan & Skeem, risk assessment tools can assist in determining whether, on utilitarian crime-control grounds, an offender should be sentenced to the upper bound of that range.\textsuperscript{77} The second role is to inform decisions regarding supervised release of lower-risk offenders.\textsuperscript{78} Monahan & Skeem note that risk assessment tools can provide an empirical estimate of whether an offender has a sufficiently low likelihood of committing an additional crime to justify an abbreviated period of incapacitation, supervised release, or no incapacitation whatsoever.\textsuperscript{79} Within a severity range set by moral concerns regarding the criminal, a risk assessment tool or algorithm can assist in determining whether, on utilitarian crime-control grounds, an offender should be sentenced to the lower bound of that range.\textsuperscript{80}

\textsuperscript{73} Christin, Rosenblat & Boyd (n 8)
\textsuperscript{74} Christin, Rosenblat & Boyd (n 8) p. 1
\textsuperscript{75} Ibid.
\textsuperscript{76} Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53) p. 493
\textsuperscript{77} John Monahan & Jennifer L. Skeem, ‘Current directions in violence risk assessment,’ 2011 Current Directions in Psychological Science 20(1) 38-42
\textsuperscript{78} Ibid.
\textsuperscript{80} Ibid.
The final role of risk assessment tools in criminal sentencing according to Monahan & Skeem is to inform decisions designed to reduce offender risk status.\textsuperscript{81} Validated risk assessment tools can be used to identify offenders who are higher-risk to prioritize for more intensive services, and adapting service plans for lower levels. According to Lowenkamp et al, programs that match the intensity of correctional services to offenders’ risk level have been shown to reduce recidivism.\textsuperscript{82} Moreover, some tools can be used to estimate an offender’s risk state, or the current likelihood of recidivism compared to his or her past likelihood.\textsuperscript{83} These tools contain variable risk factors that can be used to monitor ‘ebbs and flows’ in an offenders risk state and adjust levels of services or supervision accordingly.\textsuperscript{84} Monahan and Skeen argue that if each of these roles for risk assessment in criminal sentencing is successfully accomplished, these tools can advance the crime control objectives of the criminal law.\textsuperscript{85} The question then arises as to whether the automated data processing systems and algorithms that are being used in criminal sentencing can accurately fulfil the roles mentioned by Monahan & Skeen in a fair and just manner that aligns with individual due process rights.

\textbf{Studies on Algorithmic Tools in Sentencing}

The majority of the literature in the area of risk assessments used in criminal sentencing comes from the United States. In various states there are calls for reforming the criminal justice system, propelled by concerns about the social and financial costs of mass incarceration. There is growing concern and momentum to introduce policies and practices that avoid worsening the devastating impact on socioeconomically marginalized communities and minority groups.\textsuperscript{86} The problem is not limited to the United States, and

\textsuperscript{81} Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53) p. 494
\textsuperscript{82} Christopher T. Lowenkamp, Edward J. Latessa & Alexander M. Holsinger, ‘The Risk Principle in Action: What Have We Learned From 13,676 Offenders and 97 Correctional Programs,’ 2006 \textit{Crime and Delinquency} 52(1) 77-93
\textsuperscript{83} Jennifer L. Skeem & Edward P. Mulvey, ‘Monitoring the violence potential of mentally disordered offenders being treated in the community,’ in Care of the Mentally Disordered Offender in the Community (Oxford University Press 2002)
\textsuperscript{84} Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53). p. 494
\textsuperscript{85} Ibid.
\textsuperscript{86} Eijk (n 26) p. 475
although the call to reform sentencing practices in other countries may be less urgent, risk assessment tools are welcomed as a way to cope with diminished resources and the ongoing quest to protect the public from offenders.\textsuperscript{87} Before delving into those American studies and case law it is important to highlight examples from other regions where the popularity of the use of risk assessment tools and algorithms in sentencing is growing. Raynor and Lewis analyzed British studies in which risk-assessed offenders have received similar disposals, and they found that in almost all cases, minority ethnic offenders have been assessed as presenting, on average, a lower score than Caucasian British offenders on various risk-need instruments.\textsuperscript{88} In most studies, this difference is significant and they found that this suggests that members of minority ethnic groups are likely to receive, on average, more severe sentences than Caucasian offenders assessed as presenting similar levels of risk and criminogenic need.\textsuperscript{89}

A study conducted by Carr & Maguire on the use of risk assessment tools by judges in Ireland found that judges tended to welcome sentencing recommendations from risk assessment tools included in pre-sentencing reports (PSR), however, the lack of legislative guidance surrounding when and for what purposes they should be used raised issues regarding equality in sentencing and timely sentencing.\textsuperscript{90} The authors contrasted the Republic of Ireland to Northern Ireland noting that in the North there is a presumption that the court should obtain a PSR prior to the imposition of a custodial sentence, whereas the lack of legislative guidance mandating PSR’s in the Republic has left defendants in a position of ‘adjourned supervision,’ for weeks and then their cases have been struck off following a judge receiving a PSR.\textsuperscript{91} Carr & Maguire take a unique stance as they advocate for the use of risk assessment tools and considered the implications of the lack of legislative guidance mandating risk assessment tools in the Republic, rather than assessing the due process implications of the actual risk assessment tools themselves.

\textsuperscript{87} Ibid. p. 476
\textsuperscript{88} Peter Raynor & Sam Lewis, ‘Risk-need Assessment, Sentencing and Minority Ethnic Offenders in Britain,’ 2011 \textit{The British Journal of Social Work} 41(7) 1357-1371
\textsuperscript{89} Ibid.
\textsuperscript{90} Nicola Carr & Niamh Maguire, ‘Pre-sentence reports and Individualised Justice: Consistency, Temporality and Contingency,’ 2017 \textit{Irish Probation Journal} 14 52-71
\textsuperscript{91} Ibid.
Various studies have also been conducted on risk assessment tools used to evaluate the recidivism risk of those charged with sexual offences against children in the United Kingdom. A study found that various tools including the popular Risk Matrix 2000 and the Static-99 overestimate the risk of recidivism in those for Internet offenders.  

‘Internet offenders’ meaning those who were charged with the possession of images depicting child sex abuse. Despite the fact that none of the offenders in the study had any further convictions or charges, neither of the instruments placed any of the offenders in the ‘low-risk’ category.  

Also interestingly, it was found that the offenders who were given a lower output such as ‘medium risk’ were more likely to have accessed images that are considered the most abhorrent in nature and classified as ‘level 5,’ whereas none of the ‘high-risk’ offenders were found to have collected level 5 images.  

Moreover, a greater proportion of lower-risk offenders were found to have the largest collections of child pornography, compared to the high-risk offenders. The authors noted that this could be due to the sampling population, however, it is interesting to note that the risk assessment tool appeared to produce results that were inconsistent with the facts of each offenders case. Moreover, the authors noted that ‘the implications of this study are important to sentencing practice, as allocation to treatment may be made by assessment of immediate risk on completion of the sentence.’

Now moving to the studies based out of the United States, Surya, Mattu, Kirchner & Angwin conducted an in-depth study on the COMPAS algorithm and found that the algorithm skewed towards labelling black defendants as high-risk and white defendants as low-risk despite not using ethnicity as a variable in the algorithm. The study conducted on 7000 people arrested in Florida found that the algorithm incorrectly labelled black defendants as repeat offenders almost twice as much as white defendants, and it wrongly

---

93 Ibid.  
94 Ibid.  
95 Ibid.  
96 Ibid. p. 21
labelled white offenders who went on to re-offend as low-risk almost twice as often as black re-offenders (48 per cent vs. 28 per cent). Further, the COMPAS system demonstrated unevenness and bias when predicting recidivism between genders. Women ranked high-risk recidivated 47.5 per cent of the time, whereas high-risk men recidivated 61.2 per cent of the time. A high-risk woman would therefore have a lower risk of recidivating than a high-risk man – a factor, law enforcement or judiciary may not take into account during decision-making. In contrast, Herrschaft examined the accuracy and validity of the Correctional Offender Management for Profiling Alternative Sanctions (COMPAS) tool using a diverse sample of male offenders released to parole supervision in New York City. This study indicated that COMPAS demonstrated high predictive efficacy and validity on its associated outcome, parole revocation for a technical violation, in this case. Herrschaft noted that ‘parole is considered a privilege rather than a right, making parole decisions based on inaccurate assessment, an issue of fairness and justice.’ It is important to note that Herrschaft’s study was conducted in relation to parole rather than in regards to criminal sentencing, which, as noted above, requires much more complex reasoning. It is also important to note that there are various studies that have been conducted on COMPAS and have been completed in cooperation with agencies that are already under contract with Northpointe (the developer of COMPAS), or by individuals employed by Northpointe themselves. Northpointe responded to ProPublica’s study by criticizing the results and declaring that their software is unbiased.

The COMPAS algorithm includes input variables relating to previous incarceration, drug use, convictions your parents may have obtained, as well as whether family

97 Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, “How we analyzed the COMPAS recidivism algorithm,” 2016 ProPublica
98 Ibid.
members have been victims of crimes. Further, it includes more complex and subjective socioeconomic inquiries such as high school grades, neighbourhood crime rates and even ones chances of finding work that pays above minimum wage.\(^{102}\) Once the scores on each question or variable are assessed, the calculation of risk score is mechanical meaning each possible value of each variable corresponds to a particular increase or reduction in risk estimate in all cases. Starr notes that if an algorithmic tool or automated data processing system contains gender as a variable, ‘men will always receive higher risk scores than otherwise identical women (because averaged across all cases, men have higher recidivism rates), even if the context is one in which men and women tend to have similar recidivism risks or in which women have higher risks.’\(^{103}\) Starr also argues that demographics and socioeconomic variables receive substantial weight in the automated data processing systems used in criminal sentencing.\(^{104}\)

Another example from the United States wherein socioeconomic and demographic variables are heavily relied upon would be in Missouri’s Information-Based Discretionary Sentencing System. Wolff states that ‘the pre-sentence report includes a score for each defendant on a scale from -8 to 7, where “4-7 is rated ‘good;’ 2-3 is ‘above average;’ 0-1 is average;’ -1 to -2 is ‘below average;’ and -3 to -8 is ‘poor.’’\(^{105}\) According to Wolff, an unemployed high school dropout will score three points worse than an employed high school graduate – potentially making the difference between ‘good’ and ‘average’ or between ‘average’ and ‘poor.’\(^{106}\) Similarly, a defendant under the age of twenty-two will score three points lower than a defendant over forty-five.\(^{107}\) In contrast, an offender who has had four previous misdemeanor convictions that resulted in jail time, or having previously served time in prison on a felony conviction, or having previously escaped prison only results in the subtraction of one point.\(^{108}\) Having previously had parole or probation revoked also only results in the subtraction of one point.

\(^{102}\) Starr (n 15)

\(^{103}\) Starr (n 15) p. 813

\(^{104}\) Ibid.

\(^{105}\) Michael A. Wolff, ‘Missouri’s Information-Based Discretionary Sentencing System,’ 2006 Ohio State Journal of Criminal Law 4(1) 95-120 p. 113

\(^{106}\) Ibid. p. 112-113

\(^{107}\) Ibid.

\(^{108}\) Ibid.
point using this algorithm.\textsuperscript{109} It is clear from this criminal sentencing algorithm that socioeconomic and demographic variables do receive substantial weight, meanwhile the actual offence and severity of the offence that the accused has been charged with received no weight.\textsuperscript{110}

Surprisingly, the majority of automated data processing instruments used in criminal sentencing do not actually include the crime for which the defendant was convicted in the case at hand as a variable.\textsuperscript{111} Despite the different variables and their weights in the system, a number of U.S. Courts have enthusiastically endorsed the use of these automated data processing systems in sentencing. For example, in 2010 the Supreme Court of Indiana upheld against a state-law challenge, the use in sentencing of the Level of Services Inventory-Revised (LSI-R).\textsuperscript{112} The LSI-R was developed in Canada and is the most popular prediction instrument in use among the states in the U.S. that have not adopted their own specific instruments,\textsuperscript{113} despite the fact that its manual states that ‘it was never designed to assist in establishing a just penalty.’\textsuperscript{114} The LSI-R also includes variables relating to the defendant such as ‘dependence on social assistance’ and variables that are out of their control such as if their parents had criminal backgrounds.\textsuperscript{115} The accused with these attributes would be considered higher risk than those who do not in regards to criminal sentencing. The Indiana Supreme Court found that:

‘Such risk assessment instruments enable a sentencing judge to more effectively evaluate and weigh several express statutory sentencing considerations such as criminal history, the likelihood of affirmative response

\textsuperscript{109} Ibid.
\textsuperscript{110} Ibid.
\textsuperscript{111} Starr (n 15)
\textsuperscript{112} Malenchik v. State [2010] 928 N.E.2d 564, 572-74 Indiana
\textsuperscript{113} Starr (n 15) p. 812
\textsuperscript{114} Department of Correctional Services, ‘LSI-R Training Manual 8,’ 2002 cited in Christopher D. Webster & Stephen J. Hucker ‘Violence Risk Assessment and Management’ (John Wiley & Sons Ltd 2007) p. 88
\textsuperscript{115} Starr (n 15)
to probation or short term imprisonment, and the character and attitudes indicating that a defendant is unlikely to commit another crime.’”

The Virginia and Utah Criminal Sentencing Commissions have also developed and applied risk assessment tools that are used to assess non-violent offenders bound for incarceration. A study conducted by Wandall found that the use of the risk assessment tool in sentencing in the commonwealth of Virginia challenged the fundamental quality of modern penal justice, recognizing the moral personality of individuals subjected to criminal sentencing. The methodology of this study contrasted the Commonwealth of Virginia to Denmark’s penal sentencing system as Denmark had not implemented actuarial technologies at the time, and found, through contrasting cases of similar facts that the use of actuarial technologies would increase the prison sentence for the defendant, rather than decrease prison populations as per the aforementioned goal of the adoption of these tools. Despite this, the American Bar Association has advocated for states to adopt risk assessment tools in an effort to reduce recidivism and increase public safety. The Wisconsin Supreme Court has also endorsed the use of algorithmic technology in criminal sentencing in the controversial case of the State v Loomis.

*State of Wisconsin v. Eric L. Loomis*

The State of Wisconsin v. Loomis is a landmark case on the legality of the use of risk assessments in criminal sentencing. A detailed analysis of this case is attached to this project within the appendix; however, a brief synopsis of the facts will be provided here:

---

116 Malenchik (n 112) p. 12
117 Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53).
119 Ibid. 187
120 American Bar Association, ‘State Policy Implementation Project,’ Criminal Justice Section accessed: https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_civilcitations.authcheckdam.pdf
121 State v. Loomis (n 21)
for context. Moreover, the due process issues of this case will be analyzed further within Chapter 2 of this project. Loomis was involved in a drive-by shooting and was sentenced, based partially on the risk scores of COMPAS assessment, to six years imprisonment and five years of extended supervision.\textsuperscript{122} Loomis appealed to the Wisconsin Supreme Court on the grounds that the consideration of COMPAS in sentencing was a violation of his right to due process for three reasons: (1) it violates a defendant’s right to be sentenced based on accurate information, in part because the proprietary nature of COMPAS prevents any defendant from assessing it’s accuracy; (2) it violates a defendant’s right to an individualized sentence; and (3) it improperly uses gendered assessments in sentencing.\textsuperscript{123} The Wisconsin Supreme Court affirmed the circuit court’s sentence concluding that if used properly, observing specific limitations and cautions set forth therein, circuit court’s consideration of a COMPAS risk assessment tool at sentencing does not violate a defendant’s right to due process.\textsuperscript{124} Moreover, the court argued that the use of the algorithmic tool was merely a ‘support’ tool for consideration and not determinative in sentencing. As a response to Loomis’s arguments, the Court created and applied a number of cautions and limitations to the consideration of risk assessment tools in sentencing.

The Court requires that circuit courts should be provided with a written warning that informs the court of (1) the proprietary nature of COMPAS which prevents disclosure of information relating to how risk scores are assessed and how variable inputs are weighed; (2) that the risk assessment is based on a national sample, but no cross-validation study has been conducted for a Wisconsin population; (3) to note the various studies that indicate that COMPAS disproportionately classifies minority offenders as having a higher risk of recidivism; (4) risk assessment tools must be constantly monitored and re-normed due to changing populations and subpopulations; (5) COMPAS was not developed for sentencing, but is intended for use by the Department of Corrections in making decisions regarding parole and supervision.\textsuperscript{125} There were no dissenting opinions.

\textsuperscript{122} Ibid.
\textsuperscript{123} Ibid.
\textsuperscript{124} Ibid.
\textsuperscript{125} Ibid.
in this case, however, Justice Abrahamson emphasized that the court’s lack of understanding of the COMPAS algorithm is a significant problem and criticized the Court for denying Northpointe’s motion to file an amicus brief regarding the COMPAS algorithm.

This decision may have been an attempt to put limitations on the current trend of the use of risk assessment tools in sentencing in courtrooms throughout the United States; however, many some have argued that this decision did not go far enough in protecting the right to due process. Han-Wei Liu, Ching-Fu Lin & Yu Jie Chen stated that ‘Mere written warnings do not seem to be able to adequately inform judges as effective gatekeepers, especially when they may not be sufficiently equipped with expertise as to understand the workings of such tools.’ 126 Moreover, the court ignored the phenomenon of automation bias, which will be discussed in Chapter 2.

Although Loomis is a relatively new development, it has been cited in other U.S. cases. In the State of Iowa v. Gordon, the Iowa Court of Appeal reversed and remanded the case for resentencing, as they found no statutory authority for the use of the risk assessment scores in sentencing.127 The Iowa Court stated that ‘Loomis cannot be interpreted to prohibit the sentencing court’s use of risk assessment information in ordering incarceration because Loomis affirmed the defendants incarceration sentence.’ 128 This decision was directly contrary to Loomis in that it considered the risk scores as an ‘aggravating factor’ in deciding that the defendant should be incarcerated rather than released into the community under supervision.129 The Gordon case was not decided on constitutional grounds but rather the lack of statutory grounds for the allowance of the consideration of risk assessment scores in sentencing and also the lack of existence of any cross-validation study for an Iowa population. Given this fact, and the level of the Court

---

126 Han-Wei Liu, Ching-Fu Lin & Yu Jie Chen, ‘Beyond State v Loomis: artificial intelligence, government algorithmization and accountability,’ 2019 International Journal of Law and Information Technology 27(2) 122-141 p. 9
128 Ibid.
129 Ibid.
in the *Gordon* case, its outcome may not hold as much influence or weight as the *Loomis* decision does.

**Non-American Case Law**

The United States is not the only country wherein the consideration of the constitutionality and suitability of the use of risk assessment scores in sentencing has become a Court issue. The Supreme Court of Western Australia also considered this issue in the *Director of Public Prosecutions for Western Australia v. Mangolamara*.\(^{130}\) In *Mangolamara*, the defendant appealed his 32-month prison sentence without parole under the *Dangerous Sexual Offenders Act 2006 (WA)*, as his sentence was based, in part, on a risk assessment tool that deemed him to be ‘high-risk of reoffending.’\(^{131}\) The risk assessment tools known as the Sexual Violence Risk-20 (SVR-20) and the Risk for Sexual Violence Protocol (RSVP) are twenty item tools with three categories that consider both static and dynamic factors, much like the COMPAS tool.\(^{132}\) The Supreme Court of Western Australia found that ‘the research data and methods underlying the risk assessment tools are assumed to be correct, but this has not been established by evidence.’\(^{133}\)

In considering that Mr Mangolamara came from a remote Aboriginal community, the Supreme Court of Western Australia highlighted the fact that risk assessment tools ‘were not devised for and do not necessarily take account of the social circumstances of indigenous Australians in remote communities.’ The Court concluded with having ‘grave reservations as to whether a person of the respondent’s background can be easily fitted within the categories of appraisal presently allowed for by the risk assessment tools.’\(^{134}\) ‘The careful treatment by the Supreme Court of Western Australia poses a stark contrast to the otherwise prevailing inclination to place essentially unrestrained trust with

---

\(^{130}\) *Director of Public Prosecutions for Western Australia v Mangolamara* [2007] 169 A Crim R [2007] WASC 71 [165]

\(^{131}\) Ibid.

\(^{132}\) Ibid. 90

\(^{133}\) Ibid 165

\(^{134}\) Ibid. 166
technology, which is manifested beyond the US context as well.\textsuperscript{135} ‘Without a thorough understanding of the tools and technical expertise at their disposal, judges will most likely maintain institutional inertia and shy away from vigorous debates about their problems.’\textsuperscript{136}

In Canada, Mr Jeffrey Ewart, a Metis Canadian, is currently serving two concurrent life sentences in maximum-security settings. Mr Ewart challenged the use of five risk assessment tools used by the Correctional Service of Canada to assess an offender’s psychopathy and risk of re-offending. The basis for the Constitutional challenge of these instruments was that they were developed and tested predominantly on non-Indigenous populations and that no empirical research confirmed that they were valid when applied to Indigenous persons.\textsuperscript{137} Sample questions from the instruments included questions such as: ‘What kind of neighbourhood do you live in?’ ‘Do you know anyone who is involved in crime?’ ‘Are you bothered by uncontrollable urges?’\textsuperscript{138}

Mr Ewart claimed that the reliance on these tools with respect to Indigenous offenders breached s. 24(1) of the \textit{Corrections and Conditional Release Act (CCRA)} which requires the Correctional Service of Canada to ‘take all reasonable steps to ensure that any information about an offender that it uses is accurate, up to date and complete as possible,’ as well as ss.7 of the Canadian Charter of Rights and Freedoms (Right to life, liberty and security of person) and ss. 15 (Equality before and under law and equal protection and benefit of the law).\textsuperscript{139} The appeal was allowed in part as the Correctional Service of Canada was found to be in breach of its obligations set out in s. 24 of the CCRA, however the use of the tools were not found to breach Mr Ewart’s Constitutional rights.\textsuperscript{140} The analysis by the court found that CSC’s did not fulfil its duty to ensure that any information about an offender that it uses is as accurate as possible. However, Mr Ewart had a duty to show on the balance of probabilities that the use of the tools with

\begin{itemize}
\item \textsuperscript{135} Liu, Lin, & Chen (n 126) p. 13
\item \textsuperscript{136} Ibid.
\item \textsuperscript{137} \textit{Ewart v. Canada}, 2018 SCC 30, [2018] 2 S.C.R 165
\item \textsuperscript{138} Ibid.
\item \textsuperscript{139} Ibid.
\item \textsuperscript{140} Ibid.
\end{itemize}
respect to Indigenous offenders had no rational connection to the government objective. According to the court, he had not met this obligation, as there was no evidence before the trial judges that how the impugned tools operate in the cases of Indigenous offenders is likely to be so different from how they operate in the case of non-Indigenous offenders. Moreover, the court held that there was no evidence to suggest that the tools overestimate the risk posed by Indigenous offenders or lead to harsher conditions of incarceration or the denial of rehabilitative opportunities because of that overestimation.  

The Supreme Court of Canada’s analysis, in this case, brings to light various problems that offenders may face when attempting to challenge these automated data processing tools on constitutional or due process grounds. For example, Mr Ewart could not provide any evidence that the tools would tend to overestimate the risk of Indigenous offenders, however, Court seemed to accept that there was no available empirical research conducted on these tools to allow the appeal under s. 24 of the CCRA. Was Mr Ewart supposed to conduct his own study on these tools or pay for a group of academics to conduct a study on these tools in order to have evidence that is substantive enough to have the court allow a constitutional challenge? Canadian academics have also concluded that the intensity of Indigenous marginalization makes them more likely to score higher on risk assessments than non-Indigenous offenders. However these studies were not presented as evidence in this case.

None of the case law thus far has ruled that the use of algorithmic tools in criminal sentencing is a violation of due process rights, despite the aforementioned studies that have found that these tools can produce biased or overly harsh outputs. If a court has struck down the use of the algorithmic tool at sentencing, it has been on grounds other than Constitutional or human rights grounds. This is interesting given that in both the Ewart and Mangolamara case, both the Canadian and Australian Courts did express reservations about utilizing the tools on marginalized aboriginal populations, yet neither court considered there to be human rights or equality infringements. Moreover, all

141 Ibid.
142 Kelly Hannah-Moffat, ‘Criminogenic Needs and the Transformative Risk Subject,’ 2005 Punishment & Society 7(1) 29-51
courts in the aforementioned cases addressed the fact that the use of the tools lacked empirical validity and the dissenting judge in *Loomis* outlined the serious lack of understanding of these instruments but did not deem it necessary to prohibit their use. The following section moves to a discussion and analysis on the construction of algorithms and automated data processing tools and the issues that may arise from their composition and contributing actors.

**The Construction of Automated Data Processing Systems: Contributors & Opacity**

Currently, a wide and varied range of actors contributes to the construction and implementation of algorithmic and automated data processing systems within the criminal justice system. Governmental and non-profit organizations, along with private corporations contribute to the construction of these algorithms using unique resources. For example, the Kent police force has made substantive use of a predictive policing tool developed by a private American corporation named Predpol, while the Durham constabulary makes use of their algorithmic tool named HART, which was developed by statistical experts from the University of Cambridge in collaboration with the constabulary.\(^\text{143}\) Data sets, computing skills and testing methods of predictive instruments vary depending on the technology developers themselves, and their financial means.\(^\text{144}\) The sophistication of the modeling techniques used, the size of the data set, the amount of data missing and the overall quality of the algorithm will vary depending on the financial means of the organization constructing the software.\(^\text{145}\) For example, the Arnold Foundation’s PSA pre-trial algorithm uses a database of over 1.5 million cases from 300 jurisdictions, whereas other instruments rely on data as small as several thousand cases.\(^\text{146}\) This could cause issues of reliability as inputting the variables of a defendant that is being sentenced in a small town, into an algorithmic tool that is using New York Cities population data set could produce flawed outputs. The same can be said for

---

\(^\text{143}\) Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes, ‘Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental proportionality,’’ 2018 *Information & Communications Technology Law* 27(2), 223-250

\(^\text{144}\) Christin, Rosenblat & Boyd (n 8)

\(^\text{145}\) Christin, Rosenblat & Boyd (n 8) p.4

\(^\text{146}\) Christin, Rosenblat & Boyd (n 8) p.4
inputting the defendant’s variables into an algorithm where the data set is based on a small town where the crime rate is significantly lower. The output will likely be a lower risk score than it would be if the data set were based on the population of a large city with higher rates of crime.

Computers are well suited to discovering and analyzing patterns in input-output pairs that can guide future decision-making. In contrast to human-made rules, computerized rules for decision-making are mostly generated from historical examples. Humans orchestrate a computerized rule-creation process, rather than imparting the rules directly. This can raise problems because the systems designer does not fix the decision rule directly, and, as a result, the rule cannot be directly verified. According to Kroll et al, software code is a rigid and exact description of itself: the code both describes and causes the computer’s behaviour when it runs. This is in contrast to public policies and laws are characteristically imprecise, often deliberately so. Thus, even when a well-designed algorithm or automated data processing system does assure certain properties, there will always remain some room to debate whether those assurances match the requirements of public policy. In other words, no matter how thoroughly an algorithm is designed and built, there will always be debate as to whether it is fulfilling the public policy objectives accurately.

The construction of these automated data processing systems and algorithms and the lack of transparency is where many legal scholars and defendants take issue with them. In some cases, the algorithm is built without a data set or statistical test using what is called a ‘consensus method,’ wherein judges and criminal justice specialists agree on a variable set that, in their opinion, is significant in estimating the risk of an offender. As noted above, some tools include variables such as race, which can be particularly problematic when particular races are producing higher scores on risk assessments than

147 Kroll et al. (n 22)
148 Ibid.
others. Other immutable characteristics are often used including age, gender, education, and socioeconomic status. There will be an in-depth analysis of the impacts of using these variables on due process rights in Chapter 2.

The lack of explanation about how they produced their output is also troubling. Where judges provide pages of explanation and justification of their sentences, automated data processing systems appear opaque. Although human decision-making has always been opaque to some extent, we are innately familiar with it, and so may be less troubled by opacity stemming from it.\textsuperscript{150} Defendants facing sentences that have been decided, in part, by a risk assessment tool have no meaningful way to challenge these algorithms as they can say with an expert witness for example. Oswald et al. argue that a predictive tool cannot possibly record and assess all factors that affect the output, and all these other factors that we might neglect introduce uncertainty.\textsuperscript{151} ‘While the input datasets may be comprehensible and the code is written clearly, the interplay between the two in the mechanism of the algorithm is what yields complexity (and thus opacity).’\textsuperscript{152}

The computation phase of these algorithms is what Frank Pasquale describes as the ‘black box.’\textsuperscript{153} The problem of the black box refers to the secrecy of the algorithmic process, which thwarts meaningful scrutiny of algorithms and automated decision-making process systems that have an immense impact on society.\textsuperscript{154} The Court in the \textit{Loomis} case was challenged to address the issue of opacity of the COMPAS algorithm, but argued that the accessibility of the input and output data of the algorithm by both Mr Loomis himself, and the Court was enough to allow Mr Loomis to challenge its accuracy and therefore align with his due process rights. Liu et al. argue that the Court failed to address the ‘black box’ challenge and its legal ramifications and arguably the \textit{Loomis}
decision could impede the role of the judiciary as the final gatekeeper to protect individual’s due process rights and to provide effective remedies.\footnote{155 Liu, Lin, & Chen (n 126)}

Issues such as the heterogeneity of contributors to the construction of these systems, the contrasting availability of data, and the general lack of understanding of automated data processing systems have had legal scholars calling for greater transparency for twenty years. The form which transparency should take is far from obvious. Perhaps the most obvious approach would be to disclose the source code, but the source code of computer systems is often illegible to non-experts, and often, even experts struggle to understand what a software code will do.\footnote{156 Kroll et al. (n 22) p. 639} However, in some instances, partial opacity is required in order to prevent individuals from ‘gaming the system.’\footnote{157 Ibid.} For example, in the case of algorithms used in secondary security screening at airports, or where the regulated decision is a commercial one, and the release of such information would be undesirable because it undermines the protection of consumer data.\footnote{158 Ibid.}

Further, some large corporations may be unwilling to release their source code if they are deemed to be unreliable after further empirical study, as many of these companies, including Northpointe are for-profit and have contracts with various states. For example, Northpointe has a $1,765,334 contract with the state of Wisconsin alone.\footnote{159 Ibid.} Several other states that have contracts with Northpointe to utilize COMPAS at some stage in the criminal justice process including Florida, Michigan, Wyoming and New Mexico.\footnote{160 Ibid.} The cost of access to these programs may also impact academics accessing the tools to analyze them at all.

Overall, the issues surrounding the construction of algorithms include the incredibly varied data sets, the inability to confirm whether the algorithm will align itself with the goals of sentencing and the lack of transparency in both the source codes and

\footnote{155 Liu, Lin, & Chen (n 126)}\footnote{156 Kroll et al. (n 22) p. 639}\footnote{157 Ibid.}\footnote{158 Ibid.}\footnote{159 Freeman, (n 100) p. 85}\footnote{160 Ibid.}
technological rules. Problems can also arise regarding the statistical accuracy of these automated data processing tools in general.

**Statistical Accuracy**

According to Lightbourne, the validity measurements that risk assessment algorithms rely on, called the area under the curve, relies on the ratio of false positives to false-negatives. ‘Essentially, the area under the curve indicates the likelihood that a randomly chosen observation is correctly listed as either higher probability or lower probability than the other observation.’\(^{161}\) The industry-accepted standard is ROC = .70, meaning that a defendant is only correctly classified 70 per cent of the time, while there is a 30 per cent chance that a low-risk individual could be classed as higher risk.\(^{162}\) Lightbourne argues that given this fact, and because the U.S. criminal justice system is based on the premise that ‘it is far worse to convict an innocent man than to let a guilty man go free,’\(^{163}\) basing decisions on a tool that incorrectly labels defendants 30 per cent of the time, is unsettling at best.\(^{164}\)

**Summary**

The literature reviewed thus far has indicated that the use of algorithmic risk assessments in the criminal justice system, and more specifically criminal sentencing, presents a number of issues. The theoretical justification for the use of algorithmic tools in sentencing maintains that they should be used only if they are not resulting in the defendant’s receiving harsher punishments or longer sentences than they morally deserve. Moreover, advocates hope that these tools can lessen the burden on the court system and reduce disparities in sentencing, particularly in the context of over-incarceration of marginalized and minority groups. Studies have, however, indicated that risk assessment

---

\(^{162}\) Ibid.  
\(^{163}\) *re Winship* [1970] 397 U.S. 358, 372  
\(^{164}\) Lightbourne (n 161) p. 336
tools have the chance to predict higher rates of recidivism based upon factors that are out of a defendant’s control, (including gender, age, parents criminality, etc) along with the potential to disproportionately classify minority group defendants as ‘higher risk’ compared to non-minority defendants. Moreover, various studies have indicated that these risk assessment tools have the propensity to classify defendants with a lower socioeconomic status as a higher risk than those with an average socioeconomic status. Thus, the use of these algorithmic tools can actually produce results that are the opposite of what the advocates hoped for.

Furthermore, court’s, particularly in the United States, continue to make contractual agreements with these companies to ensure access to their risk assessment tools on a yearly basis. Court’s have shown very little restraint in regards to the use of these instruments at sentencing and would rather rely on the discretion of the members of the judiciary to essentially turn a blind eye to the scores that they will be presented with about defendants.\textsuperscript{165} Courts have also demonstrated that they require empirical evidence that the automated data processing tools produce bias or discriminatory outcomes before they will accept a constitutional challenge.\textsuperscript{166} This evidence can be almost impossible for defendants to gather themselves, and the resounding lack of empirical studies conducted on these tools outside of the United States places an almost unachievable burden on defendants in other countries to produce evidence that indicates that these tools can be discriminatory or statistically inaccurate.

Given the construction of the algorithmic tools themselves, they are very opaque and many, including the judiciary, lack understanding in regards to how they process their output. These tools do not provide an explanation like a judge would in regards to how they reached their decision. The ability to access source codes and technological rules are blocked by corporate trade secrets and as of right now, and there are no mechanisms in place to review whether data sets are fair and representative of the population that the defendant is living in. Moreover, the nature of algorithm, wherein

\begin{footnotesize}
\begin{enumerate}
\item State v. Loomis (n 21)
\item Ewart v Canada (n 137)
\end{enumerate}
\end{footnotesize}
software codes are often precise and outputs are based on the input of historical data is contrary to the imprecise nature of the human decision-making process and purposely- imprecise legislation. Thus, even the most well designed algorithms may be contrary to public policy and the intention of the legislatures. Finally, Lightbourne’s analysis is a generalization of the statistical accuracy of these tools, and the results may vary depending on the tool itself, the chosen input variables, and the population of which it is applied to. Various studies have been mentioned above that indicate a higher level of statistical accuracy, and some that indicate lower levels of accuracy. However, many of these instruments have been tested for accuracy for functions other than criminal sentencing. As mentioned, many of these tools were never designed for sentencing, but rather to assist with police or parole decision, thus there is a lack of studies regarding the accuracy of their use in criminal sentencing.

Overall, the main issues that have been identified regarding the use of algorithmic tools in criminal sentencing are: (1) Discriminatory and biased outputs from these tools caused by the use of often-immutable characteristics as input variables. (2) The opacity of the algorithmic tools. These tools do not explain their output, the same way in which a judge would. The lack of access to the source code or understanding of the algorithmic rule means that defendants and members of the judiciary are left without awareness of how these tools arrive at their conclusive output or the consequences of considering the output to a defendant’s sentence. (3) The lack of empirical evidence on these tools that prove their statistical accuracy, and evidence that indicates whether they effectively achieve the purpose that they are being used for (assisting with sentences that are fair, just, and not disproportionate to the offence).

The following chapter will address, in detail, the due process issues and implications arising from the use of these issues by applying a human rights lens to the currently available literature and case law.

---

167 Aletras et al. (n 68)
168 Osborn, Elliot, Middleton & Beech (n 92)
Chapter 2: The Implications on Due Process and Human Rights

This chapter seeks to address the question: How does the use of algorithms in criminal sentencing impact due process rights? This chapter will begin with a review of the literature regarding due process rights. Although the use of algorithms in criminal sentencing and in general is a modern concern, this due process analysis will be situated in traditional and universally accepted definitions and standards of due process rights. This chapter will also provide insight into some the human rights instruments (international and jurisdictional) that define specific due process rights in order to provide a more general analysis and distinguish this project from others that have solely focused on the United States Constitution. Following this review, this Chapter will apply these definitions and understandings of due process rights to the problems that were identified in Chapter 1.

This chapter discusses various due process implications of particular rights including the right to review and verify sentencing information; the right to an individualized sentence; and the due process implications arising from the use of particular immutable characteristics as input variables in algorithmic tools. The variables that will be addressed in this chapter include race, gender, socioeconomic status, education, and employment. Following this analysis, this chapter will provide an analysis of the arguments and counterarguments for using these characteristics as input variables and conclude by answering whether the aforementioned risks to due process rights are justified? This chapter aims to provide a unique analysis of the impact of algorithms in criminal sentencing on due process rights from a human rights perspective that goes beyond the United States Constitution. Moreover, Chapter 2 will justify the need for safeguards on due process rights and provide a natural progression into Chapter 3’s discussion on the currently available safeguards.
What Are Due Process Rights?

The first conceptual recognition of due process stemmed from England in Clause 39 of the Magna Carta 1215. Clause 39 states: ‘No free man shall be seized or imprisoned, or stripped of his rights or possessions, or outlaw or exiled, or deprived of his standing in any other way, nor will we proceed with force against him, or send others to do so, except by the lawful judgment of his equals or by the law of the land.’

Much has changed since then, and when English and American law diverged, the United States incorporated a specific right to due process within their Constitution, while England did not uphold a formal conception of due process. England does, however, retain the laws of natural justice and also the constitutional concept of the rule of law; however, some argue that these do not provide for the same implied rights that are contained within the US Constitution.

Despite the differences in formal conceptions of due process on the jurisdictional level, due process rights are arguably some of the most agreed-upon human rights on an international level. Professor Kutner said:

If there are any denials of human rights, which all nations might in principle agree, violate standards of fairness, certainly arbitrary arrest (i.e., wrongful custody without colour of legal justification) is one. If individuals may be arrested or incarcerated without cause or for causes which clearly violate fundamental human rights, they do not have the most elementary fundamental freedom.

Due process rights can be divided into two subsections – substantive due process and procedural due process. Substantive due process requires the State to ensure that laws

---

do not have an unfair impact on individuals, while procedural due process requires the State to follow procedures that ensure that the law is applied equally. Framed in another way, when one’s substantive due process rights are challenged, the question is whether the State has justifiably restricted the liberty of the person detained. The procedural issue is whether the State has followed all reasonable rules in affording the detained person the ability to challenge the basis of his or her detention. Procedural due process rights are of value has they protect substantive rights, especially the right to liberty, and partially constitute the rule of the law as they restrict the arbitrary exercise of power.

Numerous human rights instruments legislate and define ‘due process’ rights. After setting standards for dignity and freedom, the Universal Declaration of Human Rights (UDHR) devotes a cluster of articles to standards for the administration of justice, including what is often known as ‘due process.’ The UDHR enshrines some fair trial rights, such as the presumption of innocence until proven guilty, in Articles 6, 7, 8 and 11, however, Article 10 is a key provision that states:

‘Every person is entitled in full equality to a fair and public hearing by an independent and impartial tribunal, in determination of his rights and obligations and of any criminal charge against him.’

Following the adoption of the UDHR, due process rights were defined in more detail in the International Covenant on Civil and Political Rights (ICCPR). The ICCPR came into force in 1976 and the Covenant compels governments to protect the rights enshrined in the treaty and to provide an effective remedy. As of April 2019, 172 countries have ratified the ICCPR including the United States and almost every country in Europe. Article 14(1) of the ICCPR establishes the right to a fair trial and Article 14(2) provides for the presumption of innocence while Article 14(3) sets out a list of minimum

---

173 Ibid.
174 Ibid.
175 Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR) art 10
fair trial guarantees which are binding in international law on those states that are a party to it. Moreover, Article 16 of the ICCPR guarantees that everyone shall have the right to recognition everywhere as a person before the law.\textsuperscript{176}

Article 14(1) of the ICCPR reads:

‘All persons shall be equal before the courts and tribunals. In determination of any criminal charge against him, or of his rights and obligations in a suit at law, everyone shall be entitled to a fair and public hearing by a competent, independent tribunal established by law…’\textsuperscript{177}

Under the ICCPR, states are bound both ‘to respect and [to] ensure’ the rights in the covenant.\textsuperscript{178} It has become common to describe those obligations by reference to the ‘respect, protect, and fulfil’ framework, which in the context of due process rights essentially captures the idea that states must avoid actively infringing on said rights, protect those rights from infringement by others, and take positive steps to support individual realization of due process rights.\textsuperscript{179} The obligations on States to take positive steps to support the individual realization of due process rights is particularly important in the context of adapting the law to advanced technological development.

Another major human rights instrument that enshrines due process rights would be the European Convention on Human Rights (ECHR). Article 6 is the primary due process protection within the ECHR and protects the right to a fair trial, along with Article 7 which holds that no person shall be held guilty of any criminal offence which

\textsuperscript{176} International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR)

\textsuperscript{177} ICCPR Article 14(1)

\textsuperscript{178} ICCPR Article 2(1)

did not constitute an offence under national or international law at the time when it was committed.\(^{180}\)

Finally, the Fifth and Sixth Amendments in the United States Constitution provide protection for due process rights. The Fifth Amendment states that no one shall be ‘deprived of life liberty or property without due process of law,’ and the Sixth Amendment a citizen a fair and speedy trial, an impartial jury, an attorney if the accused wishes to retain one, and the chance to confront and question witnesses. Moreover, the 14\(^{th}\) Amendment, known as the Equal Protection Clause, is another right entrenched in the United States Constitution that protects the right to due process. The Equal Protection Clause is located in section one of the 14\(^{th}\) Amendment and states:

All born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside. No State shall make or enforce any law, which shall abridge the privileges or immunities of citizens of the United States; not shall any State deprive any person of life, liberty, or property without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.\(^{181}\)

Although this is not an exhaustive list of human rights instruments that protect due process rights, the UDHR, ICCPR, ECHR, and the Amendments within the United States Constitution will be the focus of this chapter given the popularity of the use of algorithmic instruments in criminal trials in the United States, and the growing use of the instruments in the criminal justice field in Europe. Moreover, case law in each jurisdiction has assisted in the development of procedural and substantive due process rights. Regardless of which particular human rights instrument, article or amendment one wishes to apply, the underlying values of due process should animate the oversight of any body, or in this case technological system, that has the ability to significantly impact

\(^{180}\) Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR) Art 6 & 7
\(^{181}\) *The Constitution of the United States*, Amendment 14, Section 1.
individuals lives, given the obligations designated under the ICCPR, and the international consensus on the importance of freedom from arbitrary detention.

The underlying values of due process are transparency, accuracy, accountability, participation, and fairness.\textsuperscript{182} It has been argued that none of the core values of due process can be fulfilled without the participation of an independent adjudicator, such as a judge.\textsuperscript{183} What happens then when an independent adjudicator utilizes algorithmic technology that has the ability to produce automated decisions about individuals? Are they no longer considered an independent adjudicator? Does the use of algorithmic technology align with the core values of due process? This chapter will demonstrate the issues that arise with the use of automated decision-making systems in criminal sentencing and the impact on individual due process rights.

\textbf{The Human Rights Issues implicated by Risk Assessment Algorithms}

There are two primary issues with the use of algorithmic risk assessment tools in criminal sentencing: 1) their impact on an individual’s right to due process and, 2) the potential that the inclusion of certain variables constitutes a violation of ‘equal protection’ under the law as noted in every human rights instrument mentioned above. Both due process rights and equal protection under the law are very much related, however, it is easier to separate them to better clarify the argument.

\textbf{The Right to Verify and Review Sentencing Information}

Criminal sentencing is one of the few areas in which automated decision-making systems are making their way into and sentencing is arguably where due process rights are most often invoked, and potentially infringed. ‘Sentencing is not merely a part of the criminal justice system, but the precise point where one’s liberty is infringed.’\textsuperscript{184}

\textsuperscript{182} Martin H. Reddish & Lawrence C. Marshall, ‘Adjudicatory Independence and the Values of Procedural Due Process,’ 1986 \textit{Yale Law Journal} 95(3) 455-505
\textsuperscript{183} Ibid. p. 476
\textsuperscript{184} Lightbourne (n 161) p. 336
Although one has been convicted at sentencing, the criminal justice system still provides protections for defendants through acknowledgement of due process protections.\textsuperscript{185} New algorithmic decision-making tools have sovereignty over important aspects of individual lives and if law and due process are absent from this field, we are essentially paving the way to a new feudal order of unaccountable intermediaries.\textsuperscript{186} In other words, this section will demonstrate that algorithmic tools used in courtrooms in the United States are currently not held to the same standard of scrutiny as various other forms of evidence or testimony, thus the absence of procedural rules and regulations leaves these algorithmic tools unconstrained and a potential threat to due process rights.

One of the main problems with algorithmic decision-making in criminal sentencing is that the defendant arguably cannot verify or review the information. Article 14(5) of the ICCPR provides that “[e]veryone convicted of a crime shall have the right to his conviction and sentence reviewed by a higher tribunal according to law.”\textsuperscript{187} The right to review of a sentence is weakened by the acceptance and use of these automated decision-making systems.\textsuperscript{188} The opacity of algorithmic tools means that a defendant [might] see the input variables and then might see the concluding output, but not see or understand the process by which the algorithm arrived at its conclusion. A defendant will likely never know the technological rule or weighting of the variables that were entered into the algorithm to produce the outcome. For example, if gender was used as an input variable and regardless of the other variables that are input, one’s gender is 90 per cent determinative of their recidivism risk output, the defendant (and judge for that matter) would likely never know this. As will be discussed in this chapter, scoring recidivism risk solely on one’s gender is arguably discriminatory and an infringement of one’s rights to equality under the law, but due to the ‘black box’ nature and construction of the algorithm, the defendant is left unaware and unable to challenge, review or verify the use of the tool.

\textsuperscript{185} Ibid.
\textsuperscript{187} ICCPR Art 14(5)
\textsuperscript{188} Raffaele Piccolo, ‘AI in Criminal Sentencing: a risk to our human rights?’ 2018 Human Rights Bulletin Law Society of South Australia 40(11) 15-17
Given the lack of understanding of these algorithmic systems, courts, defendants and enforcement agencies are no longer able to determine whether the rules have been properly applied or whether fairness obligations have been met. The use of these systems to make decisions shifts the burden of determining whether the tools comply with legal standards, obligations and human rights to external experts or to organizations creating and deploying the software rather than the courts. ‘Many computer and data scientists, as well as software engineers, are unfamiliar with the normative politics and nuances of criminal justice data – an unfamiliarity that can lead to conceptual and methodological problems.’ Moreover, the developer’s interests do not always coincide with the social interest, and the logic behind businesses does not always coincide with the need for scientific accuracy. As previously mentioned, the contracts between algorithm developers and courts in some States are of significant monetary value, thus the compliance with human rights and legal standards may not be prioritized by developers.

Algorithmic tools have been accepted into the courtroom with little scepticism. Thus far, court cases that have permitted the use of an automated data processing system have not called for an expert to provide testimony or evidence about the tool itself. In the State v Loomis, the Court denied an amicus brief from Northpointe, the creator of COMPAS, which could have potentially provided insight into the tool itself. This is in stark contrast to how courts have treated DNA evidence in the courtroom. ‘The courts took years during the 1980s and 90s to establish and accept the scientific validity of DNA and the methods used to isolate and test DNA.’ Even now, there are concerns that

189 Ibid.
190 Kroll et al (n 22)
192 Inigo De Miguel Beriaín, ‘Does the use of risk assessments in sentences respect the right to due process? A critical analysis of the Wisconsin v. Loomis ruling’ 2018 Law, Probability and Risk 17(1) 45-53 p. 49
193 Freeman (n 100)
194 State v. Loomis (n 21)
195 Kroll et al (n 22) p. 704
some forensic methods may be receiving undeserved deference from courts and thus resulting in faulty findings of fact. Following much debate, there are now strict rules and laws regarding the permissibility of evidence in the courtroom. Defendants have the right to question those presenting forensic evidence and jurors benefit from expert testimony, both of which advance due process rights. Why then, are algorithmic tools not put under the same scrutiny and held to the same standard, particularly given the vast literature on automation bias and the impact of automated outputs on human decision-making? Is the right to verify an automated data processing system scoring your likelihood of recidivism not as important as the right to review and verify forensic evidence?

It has also been argued that the use of these systems in sentencing impacts one’s right to a fair and public hearing, particularly under Article 10 of the UDHR. The Human Rights Committee has explained that: ‘[a]ll trials in criminal matters or related to a suit at law must in principle be conducted orally and publicly. The publicity of hearings ensures the transparency of proceedings and thus provides an important safeguard for the interests of the individual and of society at large. Even in cases in which the public is excluded from the trial, the judgment, including the essential findings, evidence and legal reasoning must be made public…’ Owing to the proprietary nature of the algorithms, a person can only expect the final opinion, or risk assessment, to be publicly available. Neither the information that informed the decision, nor the process of reasoning that accounts for that decision are to be revealed to the defendant, the public, or the court. Algorithms may weaken the oral, along with the public nature of sentencing, and in turn the transparency of such proceedings. Moreover, the right to a publically available judgment, which includes evidence, essential findings, and reasoning, is similarly weakened, if not curtailed, by the use of algorithms. Not only are algorithms

196 Ibid.
197 Human Rights Committee, ‘General Comment No. 32: Article 14: Right to equality before courts and tribunals and to a fair trial,’ 2007 UN Doc CCPR/C/GC/32 90th Session p. 9
198 Piccolo (n 188)
199 Ibid.
challenging the defendant’s right to review and verify sentencing decisions, but also the transparency of their trial and the right to a public hearing.

The Right to Individualised Criminal Sentencing

Although there is no statutory right to an individualised criminal sentence in the international human rights instruments, internationally there is a consensus within national case law that individual circumstance must be taken into consideration during criminal sentencing. In the United States, the Court also emphasized the role of judges crafting individualized sentences by “drawing on information concerning every aspect of a defendant’s life” in Williams v. New York.\(^{200}\) In the case of R v. Gladue, the Canadian Supreme Court held that sentencing judges are to recognise the adverse systemic and background factors that many aboriginal Canadians face and consider all reasonable alternatives to imprisonment in light of this.\(^{201}\) In R v. Ipeelee, the Supreme Court of Canada reiterated the requirement to fully acknowledge the oppressive environment faced by aboriginal Canadians throughout their lives and the importance of sentencing courts taking this into consideration.\(^{202}\) Sentencing in Australia is also founded on the principle of ‘individualised justice’,\(^{203}\) which requires that close consideration be given to the circumstances of the offence and the offender, with those circumstances bearing upon the appropriate sentencing disposition.\(^{204}\) Even the International Criminal Court highlighted the fundamental importance of individualized sentencing in Katanga where the Court declared that it had to take into account the individual circumstances of the convicted person and the global context of the conviction when it determines its sentence.\(^{205}\)

\(^{201}\) R v. Gladue [1991] 1 S.C.R 688 (Canada)
\(^{203}\) Elias v. The Queen [2013] 248 CLR 483, 494-5 para 27 (French CJ, Hayne, Kiefel, Bell and Keane JJ.) (Australia)
\(^{205}\) The Prosecutor v. Germain Katanga [2014] ICC-01/04-01/07
Algorithms and risk prediction instruments used to assist in criminal sentencing were supposed to assist in balancing the values of individual sentencing and efficient case disposal, however, many have argued that these systems are a threat to individualised sentencing and due process. Some academics have argued that inferring individual tendencies from group statistics is unconstitutional and undermines defendant’s right to due process, while others argue that on the basis of empirical findings, statistical theory and logic, it is clear that predictions of future reoffending cannot be achieved, with any degree of confidence, in the individual case. Contrary to these views, some academics assert that group data theoretically can be, and in many areas empirically is, highly informative when making decisions about individual cases, including sentencing decisions. Ultimately, the virtues of system efficiency and the voice of the unique defendant are seen as mutually competing: the more attention given to the voice of the individual, the less efficient the process of sentencing is and vice versa. Following the aforementioned examples of national case law and case law from the International Criminal Court, one would assume that the individualization of the criminal sentence would outweigh the efficiency of the algorithmic tool.

Thus far, however, the courts have admitted that an algorithmic tool like COMPAS only provides aggregate data on recidivism risk for groups similar to the offender, but have maintained that the use of the algorithm does not infringe upon a defendant’s right to an individualized sentence as the output of COMPAS was not the sole basis for the sentence and the courts have adequate information to disagree with the assessment. There has been a greater concern in the academic literature in regards to due process challenges of the use of particular variables and immutable characteristics as input factors in the algorithms used in criminal sentencing. Disparities associated with the use of risk assessment in sentencing are also a significant concern. The following sections

\(^{206}\) Starr, (n 15)  
\(^{207}\) Stephen D. Hart, Christine Michie & David J. Cooke, ‘Precision of actuarial instruments: evaluating the ‘margins of error’ of group v. individual predictions of violence,’ 2007 British Journal of Psychiatry 190(49) 60-65  
\(^{208}\) Monahan & Skeem, ‘Risk Assessment in Criminal Sentencing’ (n 53)  
\(^{209}\) Cyrus Tata, 'Ritual Individualization': Creative Genius at Sentencing, Mitigation, and Conviction,' 2019 Journal of Law and Society 46(1) 112-140 p. 115  
\(^{210}\) State v. Loomis (n 21)
outline the problems and due process challenges that can arise from considering certain variables in risk assessments as sentencing.

Algorithmic Input Variables and the Implications on Due Process

Race as an Input Variable

If the individual seeking to enforce his or her rights is black, and the adjudicator is bigoted or racially prejudiced and would therefore never find in favour of a black person regardless of the weight of evidence, all procedural guarantees of hearing, notice, counsel, transcript and examination of witnesses are rendered irrelevant. The equal protection clauses within each human rights instrument provides protection from discrimination based on race and hold that everyone is equal before and under the law, however, automated decision-making systems are arguably not held to the same standard as human adjudicators. Virtually everyone agrees that including race as a variable in an automated decision-making system would be constitutionally impermissible. Race was included as a variable in some automated data processing systems until as late as the 1970s. Nowadays, the vast majority of automated decision-making systems, particularly those used in courtrooms, do not use race as a variable in their input. However, excluding one’s race as a variable does not necessarily mean that factors that correlate heavily with an individual’s race – serving essentially as proxies for race – are excluded from these algorithms.

The input variables, which often result out of the defendant’s answers on a questionnaire in a pre-sentencing report, can be very telling of their race, despite there being no explicit question asking what race the defendant is. O’Neil states:

---

211 Reddish & Marshall (n 182) p. 476
212 Kehl, Guo, & Kessler (n 24)
213 Starr (n 15) p. 811
214 Ibid. p. 24
It’s easy to imagine how inmates from a privileged background would answer one way and those from tough inner-city streets another. Ask a criminal who grew up in comfortable suburbs about ‘the first time you were ever involved with the police,’ and he might not have a single incident to report other than the one that brought him to prison. Young black males, by contrast, are likely to have been stopped by police dozens of times, even when they’ve done nothing wrong… So if early ‘involvement’ with the police signals recidivism, poor people and minorities look far riskier.215

Other academics have argued that recidivism risk is a ‘proxy for race,’ observing that instruments give heavy weight to criminal history which is highly correlated with race.216 Similarly, it has been argued that using criminal history, as a variable may be discriminatory based on race as one’s criminal history may be influenced by past discriminatory decision-making.217 Could human rights instruments provide protection from these seemingly neutral instruments? While critics argue that the use of seemingly neutral variables that correlate with race is troubling, it is unlikely that the use of such variables would constitute an infringement upon defendant’s due process rights.

In the United States, the standard for evaluating whether a seemingly neutral law (or in this case, the use of a facially neutral factor) violates the Equal Protection Clause based on race is Washington v. Davis.218 This case involved two African American men who sought a declaratory judgment that Test 21 (a verbal ability, vocabulary, comprehension and reading test) required by the District of Columbia Metropolitan Police Department was discriminatory under the Due Process Clause of the Fifth Amendment as it was failed disproportionately by African Americans.219 The Supreme Court held that under the Constitution’s Equal Protection Clause, ‘a law or other official

218 Washington v. Davis (1976) 426 U.S. 229
219 Ibid.
act, without regard to whether it reflects a racially discriminatory purpose, is not unconstitutional solely because it has a racially disproportionate impact. Further, a plaintiff must prove discriminatory motive on the part of the state actor to receive redress under the Constitution, not just discriminatory impact.

A citizen of the United States who is subject to an automated decision in criminal sentencing would then have the challenging task of proving that there is a discriminatory motive behind the input of the particular variables that heavily correlate with race, and only a handful of cases in the forty years since the Washington v Davis case have successfully proven racially discriminatory intent. This is a particularly challenging task for a defendant in this position given that the Supreme Court of Wisconsin has already ruled that the defendant subject to the COMPAS algorithm had the right and opportunity to review and verify the information entered into the algorithm and therefore was constitutional and within the limits of the Equal Protection Clause. Loomis’s failure to succeed in his due process claim does not, of course, foreclose this line of argument in the future.

Countries outside of the United States handle these facially neutral laws and discrimination differently. In the United Kingdom, an individual need not prove that there is a discriminatory motive behind facially neutral laws or requirements to seek redress under human rights legislation. In the British case of Essop v Home Office, the Supreme Court gave a unanimous judgment in favour of Mr Essop following his claim that the assessments required by the Home Office to progress through the ranks as an Immigration Officer amounted to indirect discrimination under s19 of the Equality Act 2010. Non-white candidates, and candidates over the age of 35, like Mr Essop, were shown to pass these assessments at 40 per cent and 37 per cent of the rate of candidates.

---

220 Ibid. para 239 - 245
221 Ibid.
223 State v. Loomis (n 21)
224 Kehl, Guo, & Kessler (n 24)
225 Essop v Home Office (UK Border Agency) [2017] UKSC 27 SC
without those characteristics, however, the cause of this disparity was unknown.\textsuperscript{226} Previously, in the context of employment, indirect discrimination under s 19 of the Equality Act 2010 required a claimant to prove that they were put at a disadvantage by the facially neutral provision, which their employer applied, and which tends to disadvantage others, who shared similar characteristics.\textsuperscript{227} Given that the reason for the disparity of assessment scores was unknown, it made it almost impossible for Mr Essop to prove that the test put him at a disadvantage. The Supreme Court held that a failure to prove causation should not prohibit a claimant from suing their employer.\textsuperscript{228}

Discrimination claims based on ‘facially neutral’ policies may also be more likely to succeed in the European Court of Justice (ECJ).\textsuperscript{229} In Feryn, the director of a company maintained a policy that they did not hire immigrants due to customer preference.\textsuperscript{230} Despite the absence of an applicant who experienced discrimination based on this policy, the ECJ held that such a public statement is discriminatory based on race because it likely dissuades qualified candidates from applying for jobs.\textsuperscript{231} Interestingly, the ECJ did not address directly the question of why employer statements about immigrants can be equated to statements about racial or ethnic origin, despite maintaining policies that allow for employers to treat citizens and non-citizens differently within the EU.\textsuperscript{232}

Regardless, these were seen as a welcome judgment in the eyes of human rights advocates and human rights lawyers. This ruling avoids placing formalistic burdens of proof on claimants, unlike the requirements placed on claimants in the United States. Hypothetically then it may be easier for a defendant in the United Kingdom or European Union states to invoke human rights protections from facially neutral variables as that heavy burden to prove causation does not lie on the claimant when challenging a practice.

\textsuperscript{226} Ibid.
\textsuperscript{227} s. 19 Equality Act 2010
\textsuperscript{228} Essop v Home Office (n 225)
\textsuperscript{230} Case-54/07, Centrum voor gelijkheid van khasen en voor racismebestrijding v. Firma Feryn NV 2008
\textsuperscript{231} Ibid.
\textsuperscript{232} Linos (n 229)
or criterion. Given that there has been no case law in the EU/UK regarding the use of algorithms in criminal sentencing, this can only be stated as an assumption or hypothesis but not as fact. Moreover, the cases mentioned above involved employment criterion, rather than input variables used in automated processing systems, however, it is fathomable that a defendant in the UK/EU would evoke their equality rights under similar circumstances.

**Gender as an Input Variable**

The existing constitutional and human rights analysis of algorithms and automated decision-making systems has narrowly focused on gender (and the hypothetical use of race) but has typically not been fully developed. In contrast to race, systems such as COMPAS and LSI-R use gender as a variable. Given that freedom from discrimination based on gender is ascribed in virtually every human rights instrument and that there is a well-established history of case law condemning discrimination based on gender, one would assume that there would be considerable concern and debate regarding the use of gender in sentencing algorithms. Yet most scholarship ignores this concern or briefly asserts that the state interests are important, and therefore, outweigh removing gender as a variable in automated decision-making systems. There is, of course, considerable statistical evidence that supports that judges treat female defendants more leniently than their male counterparts, however, it is virtually unheard of for modern judges to explicitly say that they are taking gender into account.

If an algorithmic instrument includes gender as one of its variables, ‘men will always receive higher risk scores than otherwise-identical women (because, averaged

---

233 Starr (n 15)


235 Starr (n 15)


237 Carrissa Byrne Hessick, ‘Race and Gender as Explicit Sentencing Factors,’ 2010 Gender Race & Justice 13 1-13
across all cases, men have higher recidivism rates), even if the context is one in which men and women tend to have similar recidivism risks or in which women have higher risks.238 In contrast to this, it has been shown that instruments that omit gender, such as the Post Conviction Risk Assessment (PCRA) can actually overestimate recidivism for women and ultimately translate into overly harsh sanctions for women.239 Either way, the use of the risk assessment instruments appears to pose a threat to both men and women’s due process rights. Developers of algorithmic tools are stuck in a catch-22 situation wherein the omission of gender as a variable could unjustly incriminate a female defendant but the inclusion of gender could have the same result on a male.

A substantive conception of due process and equality is more concerned with the effects of the law in reality, rather than questions of whether the law on paper makes distinctions.240 Thus, the question is not whether the inclusion of gender as a variable makes distinctions, but whether the effect of this inclusion will perpetuate disadvantage, discrimination, exclusion or oppression.241 The substantive model of due process and equality may not adopt a ‘gender-neutral’ approach to distinctions; but rather it will look more favourably on measures that promote substantive equality for previously disadvantaged groups.242 Thus, a court that takes on the substantive conception of due process and equality may be more willing to include gender as a variable in algorithmic tools given that the gender-neutral review may result in higher risk scores for women. The United States has adopted this substantive conception of due process and equality.

In the United States, the Supreme Court declared that an intermediate level of scrutiny is required for equal protection issues involving classification based on gender in *United States v. Virginia*.243 To pass an equal protection challenge, the reason for gender-

238 Starr (n 15) p. 813
240 Rory O’Connell, ‘Cinderella comes to the Ball: Article 14 and the right to non-discrimination in the ECHR,’ 2009 *The Journal of the Society of Legal Scholars* 29(2) 211-229
241 Ibid.
242 Ibid
243 *United States v Virginia* [1996] 518 U.S. 515
based classification must be ‘exceedingly persuasive,’ and the State must show that the challenged classification serves ‘important governmental objectives.’\textsuperscript{244} The justifications must also ‘be genuine and not hypothesized or invented post hoc in response to litigation.’\textsuperscript{245} Scholars have argued that given the strength of data on gender and crime – as well as the importance of the government’s interest in preventing crime – it appears that the use of gender in sentencing risk assessments would pass a constitutional challenge under the American Equal Protection Clause.\textsuperscript{246} One scholar noted:

Using gender as a risk factor for violence should have little difficulty surviving an equal protection challenge: the government’s police power objective in preventing violence in society is surely “important” … and including gender as a risk factor on an actuarial prediction instrument is “substantially related” to the accuracy with which such an instrument can forecast violence – and therefore assist in prevention. Gender differences in violence are genuine and not hypothesized. And while they may be archaic, they are not outdated.\textsuperscript{247}

This is certainly a strong argument, although the data on gender and crime is vast and strong, statistics on gender and crime cannot accurately predict an individual’s recidivism risk. United States case law also dictates that inferring an individual’s tendency from group statistics is prohibited.\textsuperscript{248} In Europe, this issue is tackled slightly differently, although risk assessment tools and algorithms have yet to be addressed directly as of yet:

‘… The advancement of gender equality is a major goal in the member States of the Council of Europe and very weighty reasons would have to be put

\begin{footnotes}
\item[244] Ibid.
\item[245] Ibid. at 533
\item[246] Rebecca Foxwell, ‘Risk Assessments and Gender for Smarter Sentencing,’ 2015 Virginia Journal of Criminal Law 3\textsuperscript{435-474}
\item[248] Starr (n 15)
\end{footnotes}
forward before such a difference of treatment could be regarded as compatible with the Convention … In particular, references to traditions, general assumptions or prevailing social attitudes in a particular country are insufficient justification for a difference in treatment on grounds of sex.”

Despite this statement, European courts have appeared to follow the substantive conception of due process and equality often justifying less harsh measures against females. In the case of *Khamtokhu and Asenchik v Russia*, the Grand Chamber was divided but ultimately found that it was permissible for States to categorically exempt women, juveniles, and the elderly from being sentenced to life in prison. Similar to the aforementioned American scholars, this judgment highlighted the differences between the criminality of men and women and also focused on the importance of advancing justice. Judge Sago forwarded that,

“the same period of imprisonment for a woman is more painful than for a man, perhaps because, technically, a woman is deprived of the possibility of giving birth to a child, and in particular raising a child. This may sound like a simple gender stereotype, although many people would argue that there are biological differences and specificities of the female brain.”

The Court also found that the justification for the difference in treatment between the applicants and certain other categories of offenders, namely to promote the principles of justice and humanity, had been legitimate. It was also satisfied that exempting certain categories of offenders from life imprisonment had been proportionate in achieving those principles. Should European Courts follow this line of reasoning with regard to algorithms in criminal sentencing, the use of gender as a variable would most likely be justified, despite the *Khamtokhu* being rife with gender stereotypes. This may be

---

249 Konstantin Markin v. Russia (Grand Chamber judgment of 22 March 2012) para 127
250 *Khamtokhu and Asenchik v Russia* (application no. 60367/08) para 7
251 Ibid.
252 European Court of Human Rights, ‘Life sentencing in Russia is not discriminatory,’ 2017 *Press Release issued by the Registrar of the Court ECHR 003*
problematic as it legitimizes views of women – such as the one expressed by Judge Sajo in his concurring opinion – as victims characterized by the “specificities of the female brain,”\textsuperscript{253} as well as demonizing men and legitimizing longer criminal sanctions for them.

In contrast, the European Court of Human Rights found that there had been a violation of Article 14 in conjunction with Article 8 in the case of \textit{Ecis v. Latvia}. This case involved a male inmate who claimed that he had not been allowed to attend his father’s funeral under a law regulating prison regimes that discriminated in favour of women.\textsuperscript{254} The Court found that women who had committed a serious crime and had received the same sentence were treated differently as men were automatically placed in the highest security category, while women went to less restrictive and partly closed prisons.\textsuperscript{255} The law provided women with the possibility of attending a funeral as those placed in the highest security category were generally prohibited from leaving to attend funerals. There had been no individual assessment of the proportionality of this prohibition and the Court found that the inmate had suffered discrimination, which violated the Convention.\textsuperscript{256}

If the courts were to follow the line of reasoning in \textit{Ecis}, it may be reasonably assumed that they would be more likely to remove gender as a variable as it statistically results in men receiving higher scores despite having committed the same offence as a woman. Academics argue that the substantive approach or ‘disadvantage test’ should be the approach taken by the courts in determining discrimination based on gender as it would offer a more just application of gender equality as the key interpretative principle of the Convention.\textsuperscript{257} This argument would essentially mean that those gender variables should be included in algorithmic tools used at sentencing. This may undermine the academics petitioning for the removal of algorithmic tools in sentencing altogether,

\textsuperscript{253} \textit{Khantokhu and Asenchik v Russia} (application no. 60367/08)
\textsuperscript{254} \textit{Ecis v. Latvia} (application no. 12879/09)
\textsuperscript{255} European Court of Human Rights, ‘Prison sentence law which prevented a male inmate from attending father’s funeral led to sexual discrimination,’ 2019 \textit{Press Release issued by the Registrar of the Court} ECHR 004
\textsuperscript{256} Ibid.
\textsuperscript{257} Ivana Radacic, ‘Gender Equality Jurisprudence of the European Court of Human Rights,’ 2008 \textit{European Journal of International Law} 19(4) 841-857
however, the European courts have not had to deal with these issues yet so this analysis is hypothetical.

**Socioeconomic Status as an Input Variable**

Relatively little attention has been given to the problem of utilizing socioeconomic status as a variable in risk assessment instruments. Widely used instruments such as the LSR-I, COMPAS, the Offender Assessment System (OASys), and the Recidive InschattingSchalen (RISc) (used in the Netherlands), which are applied to offenders, include socioeconomic factors such as employment, education, financial status and residency or accommodation. This is problematic because, put simply, socioeconomic marginality contributes to a higher risk score, which in turn could translate into more severe sentences for these individuals (e.g. incarceration instead of community sanction, or a longer sentence) compared to their more privileged counterparts. Human rights organizations have long been aware that those living in poverty tend to receive harsher sentences than those in higher classes. Not long ago, the United Nations made a statement calling for urgent action to end the disproportionate impact of the most serious government-endorsed punishment, the death penalty, on those from poorer communities:

If you are poor, the chances of being sentenced to death are immensely higher than if you are rich. There could be no greater indictment of the death penalty than the fact that in practice it is really a penalty reserved for people from lower socioeconomic groups. This turns it into a class-based form of discrimination in most countries, thus making it the equivalent of an arbitrary killing... The ICCPR makes clear that all people are entitled to the equal protection of the law without discrimination, while the UN safeguards on the use of the death penalty make clear that people must have received a fair trial, including the right to adequate legal assistance, at all stages. The disproportionate

---

258 Eijk (n 26)
259 Ibid. p. 464
impact of the death penalty on the poor shows that these international standards are being violated…

The consideration of one’s wealth or socioeconomic status in sentencing presents clear due process challenges and strays further from the foundational principle that the justice system should treat everyone equally. Moreover, it leaves those with lower socioeconomic statuses caught in a catch 22. While those who are employable and socioeconomically integrated are found to be more deserving of diversion away from harsh imprisonment sentences so that they can maintain their socioeconomic status, those who lack those attributes are found to be undeserving of this diversion and, consequentially, of the opportunity to improve their socioeconomic status. In addition when a risk assessment is legitimized based on the supposed benefits of diversion for low-risk individuals, while the costs of non-diversion for high-risk individuals (i.e. incarceration and all its consequences) are ignored, the message is reinforced that convicted individuals who are also underprivileged are less worthy of opportunities to improve their life.

The United Nations produced a report wherein a human rights framework was applied to demonstrate that, while individual or community poverty may not be a violation of human rights, often State’s actions or omissions that cause, worsen or perpetuate poverty amount to violations of human rights. In this context, penalization measures represent a serious threat to States’ observance of their human rights obligations. One could argue that the use of socioeconomic status as a variable in the context of sentencing would not only penalize those living in poverty but also perpetuate individual poverty due to the lack of opportunities and stigma associated with former imprisonment. Moreover, many human rights instruments and non-discrimination statutes

---


261 Eijk (n 26)

262 Ibid. p. 470-471


264 Ibid. p. 4
apply to certain protected characteristics such as sex or race, but do not apply to
discrimination based on socioeconomic characteristics like financial status. This leaves
defendants subject to these algorithms with little statutory basis to rely on.

**Education & Employment as Input Variables**

Understandably, the use of education and employment as a variable in risk prediction algorithms is often discussed under the same umbrella as socioeconomic status; however, little has been said about the impact of algorithms and defendants in white-collar crime cases. While those with sufficiently unfavourable demographic characteristics will never qualify as ‘low-risk,’ one can surmise that those with favourable demographic characteristics will never qualify as high-risk. White-collar crime, for example, is a crime committed by a person of respectability and high social status in the course of his occupation. A study conducted by Fredericks, McComas and Weatherby found that the likelihood of recidivism amongst those convicted of white-collar offences was significantly higher than those convicted of violent offences. Although these cases did not necessarily involve the use of an algorithmic tool to assist in judicial sentencing, if defendants of white-collar crime cases variables were entered into the likes of COMPAS or LSR-I, it is incredibly likely that the instrument would provide the judge with a “low-risk” score of recidivism. Although this may not impact the defendant’s right to a fair trial, it certainly impacts the rights of victims of white-collar crime and also undermines the concept of equality under the law and the rule of law.

---

265 Starr (n 15)
267 Ibid.
The Counterarguments in Defence of Algorithms in Criminal Sentencing

The “Advisory Tool” Argument

It is obvious as to why one might oppose the use of algorithms in sentencing by appealing to the idea that ‘basing sentences on static factors and immutable characteristics – like the defendant’s education level, socioeconomic status, or neighbourhood – risk assessment may exacerbate unwarranted and unjust disparities that are already too common in the criminal justice system and society.’ There is an argument that the use of algorithmic tools in sentencing is merely advisory and not determinative in sentencing. This argument was a key holding in the State v Loomis. While it is clear that algorithms or automated data processing systems have not yet replaced judges, this argument forgets about the phenomenon coined ‘automation bias’ or ‘automation induced complacency.’ This phenomenon refers to the misuse of automation, including uncritical reliance on the proper functioning of the algorithmic tool and failure to recognize its limitations and possible failures. There is a human tendency to trust computer-generated decisions, even when a system has not been investigated or empirically studied for the purpose of criminal sentencing.

Given the heavy caseloads of courts and the tendency of individuals to trust computer software, there is a risk that judges may over-rely on these algorithmic tools and base sentences on the outputs. Automation bias must be taken into regard when discussing the due process implications of the use of these tools at sentencing as the Wisconsin court has already demonstrated that they will not accept any form of

270 Ibid.
271 Oswald et al (n 143) p. 238
Constitutional or human rights violation challenge based on the fact that the tools are supposed to be advisory.

**The Exclusion of Input Variables at the Cost of Statistical Accuracy**

Some argue that the exclusion of immutable characteristics of defendant’s from an algorithmic input variable may be at the cost of statistical accuracy. Algorithm developers will often leave out significant statistical variables, such as race or ethnicity, which could be considered as ‘problematic’ in terms of politics and policymaking.272 ‘Retaining offensive variables incites political and ethical reproach, while simply removing them weakens statistical validity of the underlying models and diminishes the promise of evidence-based practices.’273 However, most recent studies of risk prediction instruments have typically not broken down the extent to which including socioeconomic and demographic variables improves the overall predictive power of the system.274 One exception was an American study conducted in 2009 in North Carolina, which found that eliminating race, gender and prior arrests not resulting in conviction as variables from the States’ predictive instrument would have little effect on its predictive accuracy.275 A 1987 study conducted by Petersilia and Turner found that “omitting race-correlated factors from a model to predict recidivism reduced the accuracy of the model by five to twelve points.”276 A statistical advantage that minimal would not appear to justify discrimination in criminal sentencing and the complete disregard for defendant’s due process rights.

Starr also argued that for present purposes, the Petersilia and Turner study suggests that socioeconomic and demographic variables could be excluded from

---

273 Ibid. p. 279
274 Starr (n 15)
algorithms without losing any significant predictive value and that risk prediction instruments which only include criminal history and crime characteristics may be a viable alternative.\textsuperscript{277} She explained that the ‘race-correlated factors’ in the Petersilia and Turner study actually included criminal history and crime characteristics and that once those factors were already included, adding variables such as demographics, education, marital status, substance abuse and mental health variables did not significantly improve the model’s predictive value and efficacy.\textsuperscript{278} Past conduct appears to be a generally better predictor of recidivism and future criminality than static characteristics,\textsuperscript{279} however this does not address the issue of past conduct being associated with race.

The counterarguments in support of the use of algorithmic tools at sentencing can be refuted with empirically validated evidence, while none of these tools have been empirically validated themselves for the purpose of criminal sentencing. The tools may have been empirically validated for purposes other than sentencing, such as parole, but there has yet to be a study validating these tools for their use in sentencing. Overall, the inclusion of the aforementioned input variables and the opacity of the tools resulting in the impossibility to review or verify them bring rise to various due process and human rights challenges.

\textbf{Conclusion}

This Chapter provided information on the issues that arise out of the use of algorithmic tools at criminal sentencing and the resulting implications that they have on globally accepted customs and the law of due process. These challenges included the implications on the right to review and verify sentencing; the right to an individualised sentence, and the use of immutable characteristics as input variables. This chapter provided a human rights analysis on the inclusion of variables (race, gender, socioeconomic status, education and employment) that identified the potential outcomes of the inclusion of a variable and addressed the issue of statistical accuracy and impact

\textsuperscript{277} Starr (n 15)
\textsuperscript{278} Ibid. p. 851
\textsuperscript{279} Ibid.
should the variable be removed. The United State’s courts had already addressed the inclusion of gender as a variable in risk assessment tools, but the analysis conducted on the other variables used in courtrooms outside the U.S. was hypothetical, and based on precedential case law in the area of discrimination set by those courts.

Overall, this Chapter addressed the serious due process and human rights implications that arise out of the use of algorithmic tools in criminal sentencing. As the majority of courts have yet to address this issue, it is uncertain whether international human rights law instruments and domestic constitutions will provide sufficient protection from algorithmic technologies in courtrooms. However, as previously mentioned, many of the human rights instruments and non-discrimination statutes do not apply to discrimination based on variables such as financial status. Moreover, even if the human rights instruments do apply, there is a chance that the courts may justify the use of these tools or the input variables used in them for the sake of convenience or based on public policy grounds to fulfil the objective of sentencing.

The counterarguments that advocate for the use of algorithmic tools in criminal sentencing focus heavily on the advisory nature of the tools. However, they ignore the phenomenon of automation bias and the fact that these tools can lend legitimacy to a judge’s underlying bias. Moreover, the arguments regarding the inclusion of immutable characteristics as variables for the sake of statistical accuracy can be disproved by empirical evidence. Whether one is a proponent of the use of these tools in sentencing or an opponent, there is no denying that due process requires fundamental procedural safeguards of which individuals can access when subject to a decision that could affect any right of the individual. Even an advocate for these tools would agree that they must be designed for purpose, and they should be as accurate as possible. Moreover, if these instruments prove to be detrimental to human rights or inaccurate, it is fair to say that everyone would agree that there should be access to some form of remedy or safeguard for defendants subject to them. It has been argued that data protection law can assist in
protecting individuals by filling in these gaps in human rights law.\textsuperscript{280} The following chapter will analyze the currently available safeguards, most prevalently data protection law, and evaluate its efficacy in the context of protecting due process rights.

\textsuperscript{280} Frederik Zuiderveen Borgesius, ‘Discrimination, artificial intelligence, and algorithmic decision-making,’ 2019 Council of Europe
Chapter 3: Strengths and Weaknesses of Current Safeguards

Regardless of whether academics or members of the judiciary support the use of algorithms in criminal sentencing, the literature and case law consistently calls for greater transparency. According to the Council of Europe, ‘transparency enhancement measures may not only facilitate scrutiny by the public but also by independent experts, commissions or specialized agencies, which in turn, may support efforts to promote compliance with consumer protection and human rights standards.’ One of the most challenging questions to answer is what form that transparency should take. There are currently a few safeguards available to those subject to algorithms and some may have the ability to enhance transparency and promote compliance with due process rights, however, there are strengths and weaknesses to each safeguard, particularly in the context of criminal sentencing. As mentioned in the previous chapter, data protection law can assist in filling in the gaps where human rights instruments are lacking. This chapter will provide an analysis of the latest data protection law concerning the due process challenges discussed in Chapter 2.

The first safeguard that was enforced on 25th May 2018 is the General Data Protection Regulation (GDPR) 2016/679. The GDPR replaced the 1995 Data Protection Directive. Following an analysis of the applicable GDPR provisions, Directive (EU) 2016/680, better known as the Law Enforcement Directive (LED), will be analyzed and an example of its transposition into domestic law, the Data Protection Act (DPA) 2018 will also be evaluated. As the GDPR is an EU law regulation, direct effect applies and EU member states are bound to recognise and enforce it. Processing of personal data for ‘law enforcement purposes’ is not covered by the GDPR but by the LED, which replaces the European Council Framework Decision 2008/977/JHA. As the LED is a Directive and not a Regulation, it requires implementation into domestic law. The DPA is an example of the implementation of this into domestic law.

281 Committee of Experts on Internet Intermediaries (MSI-NET), ‘Algorithms and Human Rights: study on the human rights dimensions of automated data processing techniques and possible regulatory impacts,’ 2017 Council of Europe p. 37
Given the popularity of courtroom algorithm usage in the United States, the U.S. approach to data protection law and the strengths and weaknesses of their current safeguards for due process rights will be discussed. The United States lacks a federalized or primary piece of data legislation and it would be beyond the scope of this project to provide an analysis for the hundreds of State-specific data protection statutes. Thus, this project will provide a brief analysis of their Federal Trade Commission Act and the most recently passed comprehensive data privacy protection policy called the California Consumer Privacy Act (CCPA.) This chapter will conclude by answering whether these safeguards can provide effective protection for due process rights against algorithmic harms in criminal sentencing.

The General Data Protection Regulation (GDPR)

There are several provisions introduced by the GDPR of the European Union (EU) that may be of some value. The background information regarding the GDPR is discussed in depth in the legislation report in the Appendix of this project, however, a brief background of the Regulation and its specified articles will be provided in order to evaluate its strengths and weaknesses in the context of safeguarding due process rights. The objective of the GDPR is to regulate the processing of personal data and rules relating to the free movement of personal data while protecting the fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.  

Although much of the literature and case law reviewed in this thesis has originated from the United States and other countries that are not members of the European Union, the territorial scope of the GDPR is explicitly global. Article 3(1) of the GDPR states that this Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the data subject is a natural person resident in the Union.  

of whether the processing takes place in the Union or not.\textsuperscript{283} It covers a much wider scope than any other data-related legislation, including United States privacy law.

Prior to the implementation of the GDPR, there were hopes that when put into practice, the law would create a ‘right to explanation’ wherein a user can ask for an explanation of an algorithmic decision that significantly affects them.\textsuperscript{284} ‘Transparency has long been regarded as the logical first step to getting redress and vindication of rights, familiar from institutions like due process and freedom of information, and is now being ported as a prime solution to algorithmic concerns such as unfairness and discrimination.’\textsuperscript{285} Transparency in the form of a right to explanation could arguably resolve some of the aforementioned due process challenges as it would promote heightened accountability to both the algorithm creators and the data controllers while also appearing to open the algorithmic ‘black box.’\textsuperscript{286} Specifically, it could assist in addressing the due process challenge of the inability to review and verify sentencing. In contrast, some academics argue that a right to explanation within the GDPR does not exist at all and that further steps are required to enhance the transparency and accountability of automated decision-making.\textsuperscript{287}

**The right not to be subject to ‘solely’ automated decisions**

The GDPR addresses profiling and automated decision-making. Article 4(4) defines profiling as:

---

\textsuperscript{283} GDPR, Article 3(1)
\textsuperscript{284} Andrew D. Selbst & Julia Powels, ‘Meaningful information and the right to explanation,’ 2017 *International Data Privacy Law* 7(4) 233-242
\textsuperscript{285} Lilian Edwards & Michael Veale, ‘Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy you are Looking For,’ 2017 *Duke Law and Technology Review* 16(1) 1-65 p. 22
\textsuperscript{286} Ibid
\textsuperscript{287} Sandra Wachter, Brent Mittlestadt & Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,’ 2018 *International Data Privacy Law* 7(2) 76-99
any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, locations or movement.‘

Profiling is often used to make predictions about people, using data from various sources to infer something about an individual, based on the qualities of others who appear statistically similar. In automated decision-making, profiling may or may not occur and solely automated decision-making requires no human intervention whatsoever. Article 22(1) of the GDPR provides the data subject with a right to opt-out of automated processing under certain circumstances. It states that ‘the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.’ This would allow the data subject to have the final say in the use of automated decision-making. There are certain exceptions that may apply such as the necessity for the subject to enter into or perform contracts or authorization under domestic laws in certain conditions, or the data subject’s explicit consent, however, data subjects retain several crucial rights including the right to ‘obtain human intervention’ and to ‘express his or her point of view and to contest the decision.’

Moreover, the Article 29 Data Protection Working Party notes that the human intervention required under Article 22 cannot be fabricated, and the controller must ensure that the oversight of the decision, and that someone who has the authority and competence to make that decision should carry it out. Essentially, Article 22 puts a general prohibition on solely automated decisions while giving the data subject a right to

---

288 GDPR, Article 4
290 Ibid.
291 Liu, Lin, & Chen (n 126) 19
292 GDPR, Article 22(1)
293 Article 29 Data Protection Working Party (n 289) p. 11-12
294 Ibid.
obtain human intervention on the part of the controller, to express a point of view and to contest the decision (Article 22(3)). In the context of algorithms currently utilized by members of the judiciary to aid their decisions, defendants can be rest assured that they will not be sentenced solely based on the output of an algorithm under Article 22 of the GDPR. Article 22, however, is not perfect. The scope of Article 22 is limited in that it only applies when the processing has been done by ‘solely’ automated means and various systems that have significant legal impacts on data subjects lives are not considered ‘solely’ automated, but rather semi-automated.

Some academics have suggested that Article 22 of the GDPR may be of value for defendants when applied to cases similar to the State v Loomis. Hypothetically, if one applied the GDPR to the Loomis case or other cases involving the use of algorithmic tools at sentencing, Article 22 may not be of any value whatsoever as it currently stands. Given that the court in Loomis contended that the score from the COMPAS algorithm was merely advisory, rather than the sole factor in their sentencing decision, this would appear preclude the application of Article 22, as the decision was not based ‘solely’ on automated processing. The Durham Constabulary have also argued that their HART algorithm only ‘supports decision-making for the custody officer’ and that a human will always remain in the loop. The definition of ‘solely’ under Article 22 also ignores the fact that it would be easy to introduce a nominal human into the loop, ‘rubber-stamping’ automated decisions and the phenomenon of automation bias. Moreover, there is evidence that issues (which are very prevalent in the criminal justice system) such as lack of time and resources lead systems which are intended to support human decision-makers to become de facto wholly automated.

---

295 Liu, Lin, & Chen (n 126)
298 Michael Veale & Lilian Edwards, ‘Clarity, surprises and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling,’ 2018 Computer Law & Security Review 34 (2) 298-404 p. 400
Despite the fact that the output of an algorithm used during criminal sentencing is a result of profiling and could undoubtedly have serious legal effects on the defendant, the term ‘solely’ in the definition under Article 22 prohibits a defendant from employing the aforementioned right. Arguably then it is fair to say that a defendant seeking to employ the aforementioned ‘right’ granted by Article 22 at sentencing is left without safeguards, other than the fact that they have the right not to be sentenced solely by an algorithm, but even this ignores the possibility of de facto automation in the future due to high-pressure trials, heavy caseloads and lack of resources that almost every judiciary struggles with, along with advances and prevalence of technology. Perhaps the other GDPR Articles could provide a stronger safeguard for due process rights or a right to explanation.

Notification Duties & Envisaged Consequences

Articles 13-14 create notification duties for data controllers, while Article 15 establishes a right of access for the data subjects. Article 15 requires the data controller to ‘provide the data subject with information about the envisaged consequences of processing, rather than an explanation of a particular decision.’ The data subject must invoke the right of access, whereas the notification duties are placed upon the data controllers. Articles 13 and 14 provide the data subject with the right to be informed about: (1) the existence of automated decision-making under Article 22(1) and (4); meaningful information about the logic involved; and (3) the significance and envisaged consequences of such processing. Article 12(7) also clarifies that information provided to data subjects must be provided in an intelligible, and clearly legible manner, with a meaningful overview of the intended processing.

299 Wachter, Mittlestadt, & Russel, (n 287) p. 9
300 Ibid.
301 Article 29 Data Protection Working Party (n 289) p. 11-12
302 GDPR, Article 12(7)
Applying Articles 13-15 of the GDPR to the *Loomis* case or a different case wherein the defendant is subject to an algorithm at sentencing, could potentially give *Loomis* the right to access more information regarding the algorithm, particularly the ‘envisaged consequences’ of its application which could provide the defendant with a greater explanation of the algorithm. This may assist in alleviating the challenge of the inability to review or verify the algorithmic tool’s process and output. The Wisconsin Court acknowledged the lack of transparency regarding COMPAS used in the *Loomis* case due to the proprietary nature of the algorithm, however, they argued that the list of input variables (a questionnaire completed by Loomis along with his pre-sentencing report) was accessible by both the defendant and the Court; therefore Loomis actually did have the capability of challenging the accuracy of algorithm.\(^{303}\) Ultimately, the Wisconsin court appeared to value the proprietary nature of the algorithm, over the defendant’s ability to challenge, review or verify it. Should future courts have to address the issue of algorithmic opacity in criminal sentencing, perhaps Articles 13-15 of the GDPR would assist them in understanding the logic and envisaged consequences of the algorithmic tool. This is hypothetical of course but it could be a step forward in defending the right to review and verify a sentence.

**Accurate, Adequate, Relevant and Necessary**

Article 5 of the GDPR was invoked by the Article 29 Working party which requires data accuracy at all stages in order to reduce flawed automated decision-making or profiling.\(^{304}\) Article 5 of the GDPR requires that personal data must be ‘accurate, adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.’\(^{305}\) This particular section may be helpful in relation to challenging the lawfulness of particular input variables. Hypothetically, defendants may be able to challenge why their postal code or education is a necessary variable for processing in concerning their recidivism risk. Moreover, as Article 29 requires data accuracy at all stages, ensuring accuracy of the logic of the algorithmic tool itself, through auditing or

---

\(^{303}\) *State v. Loomis* (n 21)

\(^{304}\) Article 29 Data Protection Working Party (n 289)

\(^{305}\) GDPR Article 5
empirical validity studies could protect those from receiving unwarranted ‘high-risk’ recidivism scores.

Does the GDPR address discrimination?

Many of the due process challenges addressed in chapter 2 focused on discrimination, whether it was based on gender, race, socioeconomic status, or variables that imply those immutable characteristics, and the issue of algorithmic programmers being put in a position wherein it is almost impossible to make an algorithm that is free from discrimination was also addressed. But does the GDPR specifically address the issue of discrimination within algorithmic systems? The most direct reference to discrimination in regards to the GDPR is found within its linked Recital 71. The final sentence of Recital 71 states:

‘In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organizational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimized, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject, and prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect.’

Technically, recitals in EU law do not carry legal repercussions and are said to be scene-setting in regards to the application of regulations. However, where a provision in a

regulation appears ambiguous, it may be interpreted in light of the recital.\(^\text{307}\) There certainly have been cases where a provision has been affected by a recital and others where the scope of the provision is affected.\(^\text{308}\) It is unclear whether the legislation within the GDPR would be interpreted in light of Recital 71, however, if it is, it could potentially provide a criminal defendant with grounds to challenge the use of input variables such as race. Unfortunately, Recital 71 does not include some of the input variables mentioned in Chapter 2 such as gender, socioeconomic status and education so gaps would remain.

The text of the GDPR itself does not directly address the issue of discrimination, however, the Article 29 Working Party reference their views on discrimination in a few ways. The first way is through the ‘safeguards’ provided under Article 22 wherein they recommend that data controllers ‘design ways to address prejudicial elements,’ ‘audit algorithms,’ and undertake ‘regular’ reviews to avoid discrimination based on data.\(^\text{309}\) The Article 29 Working Party also suggests that data controllers relying on legitimate interest grounds to justify profiling must strongly consider safeguards concerning ‘fairness, non-discrimination and accuracy.’\(^\text{310}\) They use the example of providing data subjects with information regarding how regularly their algorithms or profiling methods are tested to ‘ensure that they remain fair, effective and unbiased’ under Articles 13-15.\(^\text{311}\) Implementing regular reviews and audits of algorithms could be helpful to test for discrimination or bias within algorithms and perhaps provide defendants with grounds to challenge algorithmic tools, should the results of these audits be made accessible to them. Unfortunately, these are guidelines and not necessarily legally binding and may be somewhat idealist in regards to their actual implementation, but they are welcome guidance in regards to the protection of human rights.

\(^\text{308}\) See for example: Case C-244/95, P. MoskofAE v. Ethnikos Organismos Kapnou, 1997 E.C.R. 1-06441; Case C-288/97Consorzio fra i Caseifici dell'Altopiano di Asiago v.Regione Veneto, 1999 E.C.R.
\(^\text{309}\) Veale & Edwards (n 298) p. 403
\(^\text{310}\) (n 289) p. 21
\(^\text{311}\) (n 289) p. 14
The Law Enforcement Directive

As previously mentioned, the LED is a Directive originating from the GDPR. As it is not a regulation, it must be transposed into domestic law through national legislation. The intention of the LED is to address the processing of personal data by data controllers for law enforcement purposes. These purposes include the prevention, investigation, detection, or prosecution of criminal offences and the execution of criminal penalties. Thus, while the GDPR applies generally, criminal sentencing falls within the scope of the LED. The European Commission states that the LED ‘protects citizens fundamental rights to data protection whenever personal data is used by criminal law enforcement authorities for law enforcement purposes. It will, in particular, ensure that the personal data of victims, witnesses, and suspects of crime are duly protected and will facilitate cross-border cooperation in the fight against crime and terrorism.’

While much of the GDPR and LED text is essentially identical, there is a clear delineation from the GDPR. The LED requires that a ‘competent authority’ must carry out the data processing ‘for the purposes of the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.’ According to the LED, a ‘competent authority may include not only public authorities such as the judicial authorities, the police or other law-enforcement authorities but also any other body or entity entrusted by Member State law to exercise public authority and public powers for the purpose of this Directive.’ For the purpose of this project, the LED would be applicable to those inputting the data into algorithmic tools such as COMPAS. It is unclear whether the LED would apply to the judiciary as one could argue that they merely receive the output of the algorithmic tool and are not involved in the data.

314 LED para. 7
315 LED para. 11
processing. This question has not been addressed in case law yet, however, one could
argue that considering the output of an algorithm while determining the sentence of a
criminal offence would be enough to constitute the judiciary falling within the scope of
the LED. For this project, it would be beneficial for the defendants if the judiciary were
to fall within the scope of the LED as it may provide some protection against algorithmic
harms.

The LED contains stronger requirements than the GDPR in some instances to
demonstrate compliance, most notably through the logging requirement. Recital 24
requires that data controllers keep records of all categories of processing activities under
their responsibility. Recital 25 requires that logs be kept for the following processing
operations: collection, alteration, consultation, and disclosure including transfers,
combination and erasure.316 The logs must be used for the verification of the lawfulness
of processing, self-monitoring, ensuring integrity and security of the personal data, and
for criminal proceedings.317 The mention of criminal proceedings is particularly valuable
in this Recital as it could require data controllers to provide information to both the
judiciary and defendant regarding the algorithmic tool and input variables used. This
could potentially enhance the judiciaries understanding and provide the defendant with
grounds to challenge the tool. If the LED had been applicable in the Loomis case, the
court may not have denied the application for an Amicus brief by the developers of
COMPAS. Moreover, the judges would have been provided with logs and information on
COMPAS and Loomis could potentially have gained a greater insight into the tool itself,
rather than only having access to the questionnaire that he completed.

Article 6 of the GDPR and recital 8 of the LED lays out what constitutes ‘lawful
processing.’ Recital 8 states:

1. Member States shall provide for processing to be lawful only if and to the
extent that processing is necessary for the performance of a task carried out

316 LED Recital 25
317 Ibid.
by a competent authority for the purposes set out in Article 1(1) and that it is based on Union or Member State law.

2. Member State law regulating processing within the scope of this Directive shall specify at least the objectives of processing, the personal data to be processed and the purposes of the processing.  

It would likely take a court decision to define ‘necessary’ in these circumstances, but if a court did decide that processing personal data for determining risk is necessary for any of the aforementioned reasons, the GDPR’s general prohibition on processing of personal data would not provide any protection for defendants subject to algorithms at criminal sentencing. If they held that data processing in the form of algorithmic tools was unnecessary to perform the task of criminal sentencing and administering justice, then the GDPR, LED and DPA could provide individuals with a right not to be subject to these tools at sentencing. The argument that the task of criminal sentencing and the administration of justice has been conducted for hundreds of years without the need for automated data processing tools is strong. However, the courts may concede that the burden of heavy caseloads and an overburdened justice system may provide the justification that data processing is ‘necessary.’ Ultimately, it will not be clear if these instruments will provide a safeguard in the context of criminal sentencing until a legal decision is made as to whether this type of processing is lawful.

Should the use of these instruments at sentencing be deemed lawful, the GDPR and LED both contain provisions applying a general prohibition on the processing of ‘sensitive personal data.’ These provisions could potentially assist in a defendant challenging the use of personal data, such as race, as an input variable. However, while the LED generally follows the requirements of the GDPR, the LED does make some allowances for the operational needs of law enforcement agencies by providing some derogation from the rights of data subjects. These derogations will be analyzed in the

---

318 LED Recital 8
319 Articles 9 & 10 (relating to criminal convictions and offences) of the GDPR and Recital 38 & 39 of LED
context of the LED’s domestic implementation through the UK’s Data Protection Act 2018. This will allow for enhanced scrutiny and an insight into how the LED and GDPR are applied domestically.

The Data Protection Act 2018

Part 3 of the Data Protection Act 2018 (DPA 2018) implements the LED in the United Kingdom and applies specifically to authorities processing for law enforcement purposes. Law enforcement purposes are defined under section 31 of the DPA 2018 as: ‘The prevention, investigation or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.’\textsuperscript{320} While the GDPR puts a general prohibition on the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic or biometric data for the purpose of uniquely identifying a natural person, health-related data, sex life or sexual orientation,\textsuperscript{321} the DPA 2018 legislates certain derogations and conditions that must be fulfilled in order to permit processing of sensitive data.

Schedule 8 of the DPA lays out the conditions required for sensitive processing and only one condition must be satisfied, otherwise, the data subject must provide consent. These conditions include ‘processing for judicial or statutory purposes for reasons of substantial public interest,’ ‘processing necessary for archiving, research or statistical purposes,’ ‘processing necessary for when a court acts in its judicial capacity,’ or ‘processing necessary for the administration of justice.’\textsuperscript{322} Arguably, it would be quite easy to satisfy any one of the aforementioned conditions when it comes to processing data for the purposes of sentencing a defendant. Not only is this court acting in its judicial capacity, but also the use of these algorithms is arguably advantageous based on cost and

\textsuperscript{320} Data Protection Act 2018 c.12 Part 3
\textsuperscript{321} GDPR Article 9(1)
\textsuperscript{322} Directive (EU) 2016/680 of the European Parliament and of the Council (Schedule 8)
Are the GDPR, LED, and DPA adequate safeguards?

Overall these instruments are a step in the right direction. The legislatures are demonstrating awareness of the sensitive nature of data processing in the context of law enforcement. These instruments are currently not sufficient enough to fill the gaps in safeguarding due process rights. The legal interpretation of some sections of these instruments could potentially provide for some protections in regards to due process. For example, Article 22 of the GDPR may not provide for a right to not be subject to automated data processing tools at sentencing due to judicial intervention, but depending on the legal interpretation of ‘necessary,’ under Recital 8 of the LED, automated data processing may not be deemed necessary to perform the task of criminal sentencing. If the use of these tools is deemed necessary for performing the task of criminal sentencing, the prohibition on processing personal sensitive data may assist defendants in challenging the use of particular input variables. As already mention, the conditions under Schedule 8 of the DPA appear to be too broad and easily satisfied to provide adequate protection.

Although evaluating the impact of these instruments on one’s right to privacy, or right to a private life is beyond the scope of this project, the GDPR, LED and DPA appear to focus on safeguarding those rights more so than other rights including the right to due process. Although programs such as COMPAS are primarily used in courtrooms in the United States, this is not to say that other countries should not be prepared for this potential eventuality. Given the vast literature on the discriminatory outputs of algorithms and cases such as Loomis entering US courtrooms, it seems negligent not to address the due process challenges of algorithms in legislation like the GDPR and LED. These instruments do offer some value in that it does require that input data is accurate and

---

perhaps encourages audits of algorithms, and it could assist defendants in gaining an understanding of the ‘envisaged consequences’ of the processing, but it won’t necessarily give them a right to opt-out of the processing. This is true unless it is decided that they are lawfully unnecessary for the task of criminal sentencing. It is difficult to answer the questions which have yet to be addressed under this modern legislation, but it is hoped that the courts will approach the use of automated data processing tools in criminal sentencing with great caution.

**United States Data Protection**

There is no law akin to the GDPR in United States law. There is no primary data protection legislation in the United States, but rather hundreds of laws enacted on both state and federal levels. The only federal data protection instrument is in the form of the Federal Trade Commission (FTC), empowered by the Federal Trade Commission Act. The FTC’s principal tool is to bring enforcement actions, such as implementing comprehensive security programs, biennial assessments by independent experts, monetary redress to consumers, disgorgement of ill-gotten gains, detention of illegally obtained consumer information, and providing transparency mechanisms to consumers.\(^{324}\)

Although having a federal enforcement commission that is attempting to enhance transparency in the area of big data and algorithms, the FTC has various issues. One of the main problems is that the FTC is often forced to impose penalties on imprecise legal standards provided in a rule or order, as there is no federal data protection standard, which can constitute a denial of due process.\(^ {325}\)

It has been suggested that the United States sectoral approach to data protection is failing that and they should address the issue by implementing legislation much like the


\[^{325}\] LABMD, Inc. Petitioner v. Federal Trade Commission [2018] 894 F. 3d 1221 at 1235 (‘The imposition of penalties upon a party for violating an imprecise cease and desist order – up to $41,484 per violation or day in violation – may constitute a denial of due process’)

88
GDPR. California has just recently passed a comprehensive data privacy protection law similar to the GDPR, called the “California Consumer Privacy Act” (CCPA), which will take effect on January 1, 2020. The CCPA is similar to the GDPR in that its scope extends beyond the borders of California, they both emphasize transparency and access, however, the CCPA is primarily focused on consumer privacy rights and company disclosures. The GDPR encapsulates far more areas of law and impacts a much broader range of data controllers and subjects. The CCPA is of little assistance to defendants facing the consideration of an algorithm at sentencing, but it demonstrates the United States movement towards legislation that applies to a much larger area than just one state.

Where the United States fails in having federal data protection legislation, their Constitutional rights may provide a greater safeguard than other jurisdictions. For example, the Equal Protection Clause, which was discussed in great length in the previous chapter, provided defendants with a safeguard to challenge the constitutionality of the algorithms in criminal sentencing. Although Loomis has been one of the only cases, and its outcome did not result in a due process violation, there is still a vast future opportunity for defendants to challenge these systems on equal protection and due process grounds due to the wide range of variables that the systems utilize in processing. There is also a chance that a court in the future could permit an amicus brief from companies such as Northpointe, which was denied in the Loomis case. This could potentially provide greater insight into the technical ‘black box.’ ‘There will, of course, be major hurdles for litigants seeking to bring an equal protection challenge to these algorithms, including attributing algorithmic decisions to a state actor and overcoming proprietary protections surrounding these algorithms. However, if the courts determine that these hurdles eclipse the merits of equal protection claims, the courts will render all algorithmic decision-making immune to equal protection review.’

---

Are the current safeguards effective?

Where the EU recognizes fundamental rights that are attached to data, and therefore harms that can arise due to its use or misuse, the legal conceptualization of data in the U.S. remains unclear at best.\textsuperscript{329} While both the US and EU regulatory regimes have laws or at least mentions of prohibitions on discrimination, that may cover the use of profiling to discriminate and differentially treat data subjects based on race, religion, or other protected categories, neither the EU or US regulatory regimes compellingly account for the distinct harm caused by profiling or automated decision-making, particularly in the context of criminal sentencing.\textsuperscript{330} The GDPR is the most promising piece of legislation in regards to rights protection in data processing, but due process rights have been mostly ignored in the drafting of the GDPR and LED, despite the acknowledgement that law enforcement agencies are often processing sensitive data which can have the most serious legal outcomes, including the loss of liberty. The question of whether the automated processing will be deemed necessary for the task of criminal sentencing will have significant outcomes. If these tools were deemed necessary or ‘lawful’ as they are in the United States, one would hope that the interpretation of the LED and its domestic transpositions would prevent those from processing data related to immutable characteristics such as race. These questions must be approached with great caution in order to prevent due process and other human rights violations.

The United States sectoral approach to data protection has not been overly successful and their primary focus appears to be on the implementation of legislation that is similar to the GDPR. Their privacy laws appear to focus more on consumer-client relationships and do not encapsulate human rights that are attached to data the same way the GDPR does. However, the United States constitutional laws and Equal Protection Clause provide grounds to challenge the algorithms based on the defendant’s due process rights as has already been demonstrated. There are various hurdles and the court has

\begin{flushright}
\textsuperscript{330} Ibid. p 23
\end{flushright}
already demonstrated a failure to protect due process rights through the *Loomis* case. There is always the possibility of a new defendant challenging algorithms on due process grounds, and one would hope that the outcome would set a new precedent that protects due process rights.

**Conclusion**

The purpose of this chapter was to analyze the efficacy of various data protection legislation in protecting due process rights. There were hopes that data protection legislation could fill in the gaps where human rights law lacked in regards to algorithmic tools and their use in criminal sentencing. Chapter 3 provided an analysis of the GDPR, the LED, the United Kingdom’s DPA 2018, and a brief analysis of the United States sectoral approach to data protection. While the new European Regulation and its Directives are a step in the right direction, as it stands, there is currently not sufficient protection for those subject to algorithms at criminal sentencing. While the Recitals appear helpful in regards to the interpretation of the legislation while respecting human rights, it is too early to determine in what way these instruments will be interpreted yet. Moreover, terms such as ‘necessary’ in regards to the lawfulness of processing and particular variables of data have yet to be specifically defined in the context of automated processing at sentencing. This is unsurprising as algorithmic tools are most commonly used in courts in the United States, however, this does not mean that those interpreting this legislation should not prepare for a potential eventuality that algorithmic tools become a tool used in criminal sentencing.
Chapter 4: Recommendations & Safeguards

The use of algorithms in criminal sentencing presents various due process challenges including the right to review and verify sentencing information, the right to an individualized sentence, and the right to a sentence that is free from discrimination. It has been demonstrated that the use of static factors and immutable characteristics as input variables including racial or ethnic origin, gender, socioeconomic status, education and employment can result in a discriminatory output and ultimately impact sentencing. The previous chapter focused on the currently available safeguards for individuals subject to data processing and their strengths and weaknesses in the context of criminal sentencing. There are clear gaps and faults in the aforementioned safeguards, and the issues of opacity and discrimination in algorithms remain. International and national statutory safeguards are incredibly important as they provide individuals with transcribed grounds of which to seek remedy and defend themselves, however, Burrell has argued that a combination of approaches will be required to combat the consequences of algorithmic opacity, including context-specific regulatory frameworks.\textsuperscript{331} The difficult question then is how should we move forward?

The purpose of this chapter is to address any changes that should be made in order to better protect due process rights. This chapter will provide recommendations of safeguards that could potentially address the due process issues discussed in Chapter 2 and attempt to fill the gaps of the current safeguards that were analyzed in Chapter 3. This chapter will begin by addressing the argument of removing or putting a general prohibition on algorithms in criminal sentencing. Then, should it be found that this is not a feasible option, this chapter will outline a series of recommendations and safeguards that must be implemented in order for the use of algorithms in criminal sentencing to comply with due process rights. The recommendations will include addressing ‘semi-automated’ decision-making in legislation, implementing mandatory training for judges for assessing the reliability of the output of algorithms, requiring expert testimony on the algorithms, enhancing transparency through open source codes or explanation of rules.

\textsuperscript{331} Burrell (n 18)
regular mandatory audits, and finally the creation of algorithms with human rights in mind.

**Removal or Prohibition of Algorithms from Criminal Sentencing**

It has been argued that ‘the use of algorithmic technologies to help determine a sentence or assess the risk of re-offending might result in better informed, more consistent, and more objective judgments.’ It is difficult to determine whether judicial sentences would be more consistent with the consideration of an algorithmic output as no empirical studies have been conducted to demonstrate that fact. Further, the first chapter of this project provided various examples wherein the algorithmic output was not objective at all, but rather, discriminatory against individuals based on variables such as race. It has been argued that providing a judge with the output of an algorithm to consider in sentencing allows them to consider *all of the relevant* information about in the case in a summarized way, and leaves them better informed, however, as it stands the risk of taking inaccurate or biased information into consideration is too high, particularly when an individual’s liberty is at stake. The final argument that is often forwarded in defence of using algorithms in criminal sentencing is that they are merely advisory tools. The argument that ‘algorithm is merely one tool available to the court at the time of sentencing and a court is free to rely on portions of the assessment while rejecting other portions’ is flawed in two ways: (1) if the tool is determinative of the sentence, it is a breach of the defendant’s due process rights, or at least invokes a right to examine the evidence, and (2) if it is not determinative at all, why is it even invoked in court?

As stated in Chapter 2, states party to the ICCPR are bound ‘to respect and [to] ensure’ the rights in the covenant. It has become common to describe those obligations by reference to the ‘respect, protect, and fulfil’ framework, which in the context of due

---

332 Janneke Gerards, ‘The fundamental rights challenges of algorithms,’ 2019 *Netherlands Quarterly of Human Rights* 37(3) 205-209 p. 207
333 Christin, Rosenblat & Boyd (n 8) p. 1
334 *State v. Loomis* (n 21) p.92
335 Beriai (n 192) p. 53
process rights essentially captures the idea that states must avoid actively infringing on said rights, protect those rights from infringement by others, and take positive steps to support individual realization of due process rights.\textsuperscript{336} At present, the United States has neither avoided potentially infringing upon criminal defendant’s due process rights or taken active steps to protect due process rights through the implementation of statutory provisions or otherwise. The European Union’s implementation of the GDPR is a step forward in supporting the realization of privacy rights amongst member states but has arguably fallen short in regards to due process for defendants subject to algorithms at sentencing.\textsuperscript{337} The same can be said for the domestic implementation of the LED by member states such as the United Kingdom’s Data Protection Act 2018. Although many EU member states are not currently using algorithms in criminal sentencing, the literature and case law stemming from the United States, coupled with the growing trend of use of technology in the justice system should have come as a fair warning to legislatures of potential harms defendants may face. Both the United States and EU (and its member states) legislatures have been neglectful of their obligations to respect, protect and support the individual realization of due process rights.

In addition, the potential harm posed to an individual that could arise from disproportionately sentencing him or her based on a potentially flawed or bias algorithm is too great a risk to a fundamental human right. Moreover, the potential harm to the community from imposing light penalties that are not commensurate with the seriousness of crimes could outweigh any benefits flowing from improvements to the consistency and efficiency of the sentencing system.\textsuperscript{338} For these reasons, it is the view of this author that jurisdictions should remove algorithms from the criminal sentencing process altogether. The implementation of this ban may be through the implementation of legislation, or perhaps through the interpretation of legislation that already exists such as the GDPR or LED. For example, a court could find that the automated processing of data is not a necessary task in order to conduct the task of criminal sentencing under Recital 8 of the LED (or its domestic transposition into their countries legislation.). The United States

\textsuperscript{336} See Chapter 2
\textsuperscript{337} See Chapter 3
\textsuperscript{338} Bagaric & Wolf (n 2) p. 45
may require a ban through legislation or another case involving a due process challenge that would reach the federal Supreme Court. It is too uncertain and potentially too optimistic to depend on the courts to interpret the legislation, or wait for a case to reach the Supreme Court, thus the imposition of domestic legislation, perhaps in the form of court procedure rules, would be the best way to implement a ban on these tools in courts.

Should it be determined that algorithms become a customary part of the sentencing process, various protections must be put in place to defend the individuals due process rights. Oswald et al. proposed a concept known as ‘experimental proportionality’ designed to tackle the introduction of algorithms in the criminal justice system.\(^339\) Although these academics are based in the United Kingdom, they are also of the view that usage of algorithms will likely begin expanding there as well. Oswald et al. predict that the UK Supreme Court will take the view that, provided the interference is not substantial, and the rational connection with a legitimate aim is demonstrated, the use of data by an algorithm would be proportionate if such use is kept under a system of challenge and review that is accessible by individuals.\(^340\) Thus they have worked on creating framework’s to tackle algorithmic harms, and then applying the concept of experimental proportionality to algorithm implementation could potentially act as a safeguard.

Oswald et al’s approach to experimental proportionality is split into two formal elements. The first is a formal adoption of the approach to the doctrine of proportionality taken by the court in \(R v\) Commissioner of Police of the Metropolis and another to ‘allow the benefit of the doubt to be given to the public sector body where it is not yet possible to determine with any certainty the balance or imbalance of benefits and disadvantages in relation to the new algorithmic technology.’\(^341\) Secondly, Oswald et al suggests a change to the statutory procedure and forms of relief available so that the High Court could order that the benefits and harm risks, and so the proportionality of the particular use of the

\(^{339}\) Oswald et al (n 143)
\(^{340}\) Oswald et al (n 143) Following the case of \(R v\) Commissioner of Police of the Metropolis and another [2015] UKSC 9 p. 243
\(^{341}\) Oswald et al (n 143) p. 243
algorithm, be reviewed in another hearing after some time. They explain that this approach to experimental proportionality is one that a potential independent regulator could take when assessing algorithms, and also could inform methods of internal assessments by public bodies such as specified trial and review periods, and should proportionality not be demonstrated, further use of the technology would not be authorized.\footnote{Ibid. p. 244} Having an independent regulator or expert is also a safeguard that will be discussed further in this chapter. An expert would have to be independent from the developer company to avoid conflicts of interest.

Having the ability to ‘test-drive’ the algorithms before considering implementation is beneficial in a general sense as it allows for assessments and trial and review periods. This, however, may be a dangerous game when it comes to criminal sentencing given the potential conflicts with due process rights. A suggestion may be to conduct these trials and reviews on hypothetical actors (ie. in a mock court), rather than on an actual criminal defendant, or to assess the literature and case law arising from jurisdictions that are already utilizing these tools in their sentencing process (ex. the United States). This could be conducted in coordination with an independent expert and members of the judiciary. To concur with Oswald et al., if an algorithm is found to be proportionate to be used in criminal sentencing, a change to statutory procedure must be made to provide relief and rights to defendants.\footnote{Ibid. p. 244} The following are suggestions of safeguards for protecting defendants due process rights.

**Implementation of Mandatory Judicial Training on Assessing Algorithms**

A safeguard that must be implemented if using algorithms is to become a standardized practice in criminal sentencing is the mandatory implementation of training programs for judges to enhance their understanding of algorithms. The concerns around opacity of algorithms raise challenges for an individual’s ability to understand, and therefore question or challenge the process, and judges are not exempt from these
challenges. Understanding algorithms does require in-depth knowledge of mathematics and even computer science to comprehend how they work. Most judges have not received any formal training in the scientific method, thus there is good reason to suspect that legal decision-makers are no better at methodological and statistical reasoning than laypeople. Automation bias is the primary concern here given the risk of a judge being swayed by an algorithm, or members of a judiciary essentially becoming ‘rubber stampers’ of automated decisions.

First, members of judiciaries should be made aware that algorithms and risk assessment tools are not fallible. Judiciaries should be provided with the information arising from studies that have found algorithmic tools to be discriminatory in nature such as ProPublica’s study on COMPAS to demonstrate the fallibility. Moreover, explicit training should be provided to judiciaries on the phenomenon of automation bias. Studies have demonstrated that those provided with training on the phenomenon of automation bias are more likely to scrutinize automated systems suggestions and less likely to be complacent in accepting their outputs.

Training programs of this kind are not a new concept for judiciaries. In the United States, federal district court judges receive training in scientific theory and methodology to enhance their ability to assess the reliability of expert testimony. In England and Wales, the Judicial College offers training programs in various subjects including a

---

346 Veale & Edwards (n 298) p. 400
347 Jeff Larson, Surya Matthew, Lauren Kirchner & Julia Angwin, ‘Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks.’ 2016 ProPublica
course in ‘Assessing Evidence.’ With so much scepticism around other forms of evidence such as forensic evidence provided by scientists, it should not require much justification to implement similar training programs or workshops in relation to algorithms and automation bias.

Finally, requiring that a judiciary undergo mandatory training in assessing the reliability of algorithms would lessen the burden of opacity in sentencing and increase awareness on an issue that will undoubtedly present itself to judiciaries in the future for reasons other than sentencing. For example, a defendant may present themselves to a judiciary who has been subject to an algorithm in the policing context, or perhaps in the context of airport security and travel. Another example could be an individual claiming a breach of the recent data protection legislation such as the Data Protection Act 2018. A judge who has at least some background knowledge and has been trained in assessing the reliability of algorithms would arguably be able to find an appropriate remedy or just sentence for those subject to algorithms better than a judge who is less informed.

Overall, the benefit that could come from implementing mandatory training programs for judges in assessing the reliability of algorithms includes mitigating the effects of automation bias and rubber-stamping, enhancing transparency in sentencing, and enhancing judicial knowledge for sentencing or remedial purposes. Moreover, although not necessarily a judicial training exercise, consideration should be given to the requirement that a judge should explain their reliance on an automated systems decision in detail. Judges have a moral obligation - and in some jurisdictions a legal obligation – to provide reasons for their decisions as it is considered fundamental to the legitimacy of democracy. Therefore, asking judges to identify the extent to which they relied on the algorithmic output is not only necessary in a democracy, but it also would further mitigate the effects of automation bias and ‘rubber-stamping.’ This requirement for

judicial identification of reliance on the tools output, coupled with the training programs could assist in achieving the ultimate goal of safeguarding individual due process rights.

Expert Testimony on Algorithms

Despite the black box dangers of algorithms used in sentencing, there are currently no testimonial safeguards for judiciaries (or juries for that matter) to avail of in the context of sentencing. As it stands, there is no way to assess the credibility of algorithmic tools as an actor in the court process. The opposite is true for other forms of evidence. For example, ‘when a litigant offers a human assertion for their testimony, the law subjects it to testimonial safeguards – such as impeachment and the hearsay rule – to give juries the context to assess its credibility.’ This is why the provision of an expert on algorithms in the courtroom would be necessary. The aforementioned training programs for judges could increase their knowledge and enhance their ability to assess the reliability of algorithmic evidence, judges are not computer scientists and it would be excessive to expect a judiciary to have a completely comprehensive understanding of how algorithms function. Thus, providing expert testimony for algorithmic programs would be of benefit to both the judiciary and defendant in the context of due process rights.

Human rights instruments, including the UDHR, provide defendants in criminal trials with the right to examination of witnesses. Given that a defendant or their counsel cannot cross-examine a machine, providing the court with an expert witness on the algorithm itself would be a progressive step in realizing the defendant’s due process rights. There is a clear advantage to providing a human expert to testify the reliability of the algorithm or answer questions about the output, versus the approach taken in Loomis wherein the court determined that access to the questionnaire answers that were entered into the algorithm were enough to satisfy transparency requirements.

352 See Chapter 1
354 Article 64 UDHR
Moreover, it would provide a judiciary with the information required to determine whether they wish to consider the algorithmic output as evidence or whether they believe that it should be removed from the trial. For example, if an expert were to determine that the algorithm has a tendency to produce biased outcomes, a judge would then have the choice to remove the output as evidence for consideration. Further, providing an expert witness may also facilitate corrections of the system’s code by identifying software flaws or invalid embedded rules in the system. For example, the New Jersey Supreme Court required modifications be made to breathalyzer testing equipment following an expert review of the equipment wherein the expert witness testified that although the breathalyzer was ‘generally scientifically reliable,’ its software had several ‘mechanical and technical shortcomings.’ The Court ordered these requirements in an attempt to safeguard defendants from misleadingly high readings, which ultimately lead to the improvement of the software overall. There is no reason why this practice could not be applied to algorithms used in criminal sentencing.

An important question that arises out of the proposition of providing an expert witness on algorithms in court is who that expert should be. To best safeguard due process rights, the expert should be an independent individual who has an extensive background in algorithmic technology and computer science, but one who does not have a corporate or financial interest in the algorithmic technology. Therefore a member of staff from Northpointe should not be one to provide expert testimony at a trial wherein COMPAS is the algorithm being utilized. The technological opacity of algorithms, coupled with personal self-interest and corporate incentives arising out of the mass implementation of these algorithms in the justice system could encourage developers to provide sugarcoated or biased testimony. Thus an independent individual with background expertise in algorithms and computer science would be the better option to avoid bias. Finding and hiring an individual who is proficient in the study of algorithms

355 Citron (n 323)
356 New Jersey v. Chun [2008] N.J. 943 A.2d 114, p. 120-121
357 Kenneth A. Bamberger, ‘Technologies of Compliance: Risk and Regulation in a Digital Age,’ 2010 Texas Law Review 88(4) 669-739 (discusses the impact of corporate incentives and self-interest leading programmers to inserting forms of automation bias into their machines and the ease of hiding these biases due to opacity).
and can unravel the algorithmic decision to a court would be expensive.\(^\text{358}\) However, the enhanced due process protection, along with the opportunity to improve algorithmic programs could save money in the long run by avoiding further trial and litigation on due process infringements caused by algorithms. Furthermore, these experts could provide not only the judiciary and defendant with an understanding of algorithms, but they could enhance technological transparency on a societal level due to the public nature of proceedings and their ability to provide an explanation on the ‘rules’ underlying in the ‘black box.’

**Technological Transparency: Open Source Codes & Rule Explanations**

The right of defendants to observe and assess the accuracy of information used during sentencing conflicts with the general nature of algorithms.\(^\text{359}\) Proprietary algorithms are not easily accessible; rather they obstruct individuals outside of developers and employees of the company from examining their source code and how scores are ultimately calculated.\(^\text{360}\) To safeguard due process rights, algorithmic and technological transparency must be increased.

Each defendant should have access to all data pertaining to the data subject, and ideally, the logic of algorithms should be open to public inspection as well.\(^\text{361}\) This would open up the technological black box, providing the defendant and their counsel with grounds upon which to challenge the algorithm. Moreover, it would prevent particularly discriminatory algorithms that place heavy weighting on variables such as race from being permitted in the criminal justice system in the first place. The creator companies of these algorithms are often averse to the concept of transparent source codes and software rules. The Alan Turing Institute reported that the two largest hurdles to the ‘right of

---

358 Citron (n 323) p. 1284  
359 Freeman (n 100)  
360 Ibid.  
361 Citron & Pasquale (n 186) p. 26
explanation’ are trade secrets and copyright concerns.\textsuperscript{362} There are various policy justifications for withholding information in the case of automated decision-making including revealing trade secrets, violating privacy, or leading to gaming of the decision-making process.\textsuperscript{363}

The ‘gaming the system’ argument refers to the fear that those who understand the algorithms rules or have access to its source code could use this understanding to avoid a potentially negative output from the system. For example, algorithms are becoming increasingly popular in the airport security field for assessing threats both from the contents of traveller’s baggage and from the traveller themselves. There is a general argument against releasing source codes or rules as those wishing to exploit the system may adjust their behaviour in order to avoid being assessed as a threat in the airport security process. Another example that is often used to back this argument is the algorithm or process used to determine which tax returns to audit.\textsuperscript{364} Those who are concerned about being audited may alter their behaviour or complete their tax returns in a way that would allow them to evade an audit under the algorithm if they were able to understand the rule. These concerns are valid in these instances, however, in the context of criminal sentencing, this argument is not strong enough to justify opacity.

If a defendant were able to understand the rule in the algorithm, it would better permit him/her or their counsel to question and challenge the algorithm. Further, if the defendant were to attempt to ‘game the system,’ they would either be attempting to change immutable characteristics about themselves or commit less crime as the algorithms in criminal sentencing tend to use criminal history and immutable characteristics as input variables. It is impossible to change immutable characteristics about oneself and to retroactively commit less crime, thus if anything, a defendant understanding the rule would be an incentive for them to ensure that they do not commit


\textsuperscript{363} Kroll et al. (n 22)

\textsuperscript{364} Ibid.
further crimes while awaiting sentencing. One could also rebut releasing source codes or algorithm rules by stating that defence lawyers understanding of the rules could lead to a defendant or the defence ‘gaming the system.’ However, the same is not said about defence lawyers having access to the methods used in the collection and analysis of forensic evidence. Defence lawyers have a right to interrogate and challenge a forensic scientist who is on the stand on behalf of the defendant. There is no reason why the same should not be accepted for algorithms in criminal sentencing.

The argument for preventing open disclosure of source codes or algorithmic rules based on protecting trade secrets delegates corporate companies a disproportionate amount of power. The justice system in a democratic state is supposed to be built on the values of openness and transparency. If trade secret exemptions apply here, ‘the refusal to produce the code of automated decision-making systems allows agencies to enforce laws that no one can see or monitor, which is the very antithesis of open government and democracy.’

In the State v Loomis, Loomis did, in fact, argue that his inability to access the source code of COMPAS was a violation of his due process rights. The Court in Loomis did acknowledge the lack of transparency due to the proprietary nature and trade secrets of COMPAS but accepted the output based on the fact that Loomis had access to the input variables (answers to the questionnaire). In my view, this undermines due process and was the wrong determination due to the fact that Loomis would never know which variables carried what weight, or how COMPAS arrived at its conclusion. The dissenting justice also focused on how the lack of understanding of COMPAS was a significant problem.

Many have argued that releasing the source code is unnecessary as revealing the code to an individual would likely not assist them in understanding the algorithm.

\[\text{365} \text{Citron} \ (\text{n} \ 323) \ p. \ 1293\]
\[\text{366} \text{State v. Loomis} \ (\text{n} \ 21)\]
\[\text{367} \text{State v. Loomis} \ (\text{n} \ 21)\]
\[\text{368} \text{Ibid. para} \ 133\]
anymore than if it had not been released.\textsuperscript{369} This is why the previous safeguard suggested would provide for an expert witness in court. Moreover, the supporting guidelines for Article 15 of the GDPR in states that ‘instead of providing a complex mathematical explanation about how algorithms or machine-learning work, the controller should consider using clear and comprehensive ways to deliver the information to the data subject, for example: the categories of data that have been or will be used the in profiling or decision-making process; why these categories are considered pertinent; how any profile used in the automated decision-making process is built, including statistics used for analysis; why this profile is relevant to the automated decision-making process; and how it is used for a decision concerning the data subject.’\textsuperscript{370} The explanation should be accessible and understandable to defendants of any educational background, and should assistance be required in order to better understand the algorithm, it should be given. Perhaps there should be a written explanation, accompanied by a verbal explanation from the expert witness. This also ensures that a formal paper trail is available to members of the judiciary, prosecution and the defence, which are often required when exchanging evidentiary information. This could provide a helpful guideline in providing laymen with the source code and algorithm rule information in a non-technical manner.

**Regular Audits**

Should there be a blatant refusal to disclose source codes or algorithmic rules, there should at the very minimum be regular auditing in place to safeguard due process rights. A validation program or audit must be established that would work with the developers in assessing and monitoring the validity of the algorithm. In computer science, auditing is defined as ‘an independent evaluation of conformance of software products and processes applicable to regulations, standards, guidelines, plans, specifications, and procedures.’\textsuperscript{371} The Oxford Internet Institute explained how an audit can create a

---

\textsuperscript{369} Kroll et al. (n 22)

\textsuperscript{370} Article 29 Data Protection Working Party (n 289) supra note 85, at 31

'procedural record to [...] help data controllers to meet accountability requirements by detecting when decisions harm individuals and groups, by explaining how they occurred, and under what conditions they may occur again.'\textsuperscript{372} The Article 29 Working Party for the GDPR also recommended the use of regular auditing to test the efficacy of algorithms or automated data processing systems. In the context of criminal sentencing, it may be best practice to have scheduled audits implemented statutorily to ensure compliance and protection of due process rights.

Once again, it would be beneficial for the auditor to be independent from the company that has developed the algorithmic program to avoid bias or conflicts of interest. Moreover, a validation program would be more effective if the auditor had access to the source code, however, this may provide a compromise for companies wishing to maintain the proprietary nature of the algorithm as they could choose only to provide the source code to an impartial auditing agent.\textsuperscript{373} Validation studies should be conducted on any algorithm being used in criminal sentencing and it may also be beneficial to conduct the study in the context of the jurisdiction. The population and profile of a population vary greatly by jurisdiction; so inputting the variables of someone from a small town into an algorithm that’s data set is based on the population of London for example, may produce inaccurate results.

Overall, audits are a compromise regarding the removal of algorithms in criminal sentencing. However, they are necessary to ensure accuracy and protect due process rights. Moreover, the use of independent auditors and the implementation of a statutory framework that demands regular and scheduled audits would provide an added layer of protection for due process rights.

\textsuperscript{373} Freeman (n 100) p. 102
Designing Algorithms with Due Process and Criminal Sentencing in Mind

There are particular questions for which it is provably impossible to create an algorithm that will always give a correct answer. This is due to the nature of algorithms and because at the core of computer science is 'mathematical reasoning.' The question of recidivism risk amongst criminal defendants is arguably one of those questions. One could argue that a human judge cannot provide the correct answer every time either, however, judges do not face the same catch 22 problem that an algorithm does. The catch-22 problem is if the algorithm uses variables such as race, it will naturally be deemed discriminatory, but if those variables are excluded from the input, the algorithm can lack accuracy and result in discrimination regardless. Moreover, using gender as a variable can lead to higher risk scores for men, but excluding gender can lead to higher risk scores for women. A judge would be considered discriminatory for taking the race of the defendant into consideration, but excluding race would not lead them to make an inaccurate judgment.

A judge does not face the same technical problems that an algorithm would. For example, if a developer was to create an algorithm or software program that's prohibited output is unlawful discrimination, there is always a chance that even the most high-quality and well-written software can do something that it was not intended to do. ‘Software testing has improved such that today many bugs are caught, but it is still impossible to catch all bugs after a program is written.’ Even after audits of algorithms of this sort, the question of whether an offender is likely to commit an offence again is a difficult question to measure, even for trained judges. There are always exceptions to the rule in regards to criminality. A career criminal’s circumstances may lead them to

---

375 Ibid.
376 See Chapter 2
377 Desai & Kroll (n 374) p. 32
378 Ibid. p. 32
change their ways, and vice versa. Given the impossibility of creating a perfect algorithm for criminal sentencing, the only solution may be to take due process and human rights into account while creating the algorithm, and also when making a decision based on its output.

Thus, this final safeguard suggestion is a preventative measure rather than a reactive measure. This safeguard would involve cooperation on behalf of computer scientists and developers and require them to understand and apply due process and other human rights instruments and perspectives to the algorithms during their creation. As mentioned before, it is essentially impossible to create an algorithm that will always be accurate, however, there are steps that developers could take to create programs that are better suited to criminal sentencing. As previously stated, ‘many computer and data scientists, as well as software engineers, are unfamiliar with the normative politics and nuances of criminal justice data which can lead to conceptual and methodological problems.’

Moreover, the developer’s interests do not always coincide with the social interest, and business logics do not always coincide with the need for scientific accuracy. Thus the designers of algorithms that are going to be used for criminal sentencing are going to need to have at least some understanding of the criminal justice system and the importance of human rights for criminal defendants. Computer scientists would be required to critically assess the likely impacts and consequences of risk assessments based on the values and incentives of the system in which they are embedded.

ALGO-CARE

One potential method for restricting the discretion of algorithm creators and developers without requiring legislation containing specifications would be to require

379 Hannah-Moffat (n 191) p. 460
380 Beriaín (n 192) p. 49
381 Ben Green, ‘“Fair” Risk Assessments: A Precarious Approach for Criminal Justice Reform,’ 2018 Harvard University In 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018)
administrative agencies to publish guidance for software development.\textsuperscript{382} Administrative agencies could work together with developers to identify the properties they want an algorithm to possess, and the program could be designed to satisfy those properties and permit proof.\textsuperscript{383}

Oswald et al. designed a framework, which may come of assistance to those creating the algorithms and the judicial decision-makers. These concepts were designed in response to their analysis of the HART algorithm, used by the Durham Constabulary to determine risk and recidivism in offenders, and an academic review of the major issues faced by society from the rise of algorithms, and their impact on human rights.\textsuperscript{384} Although HART was designed for police use,\textsuperscript{385} the underlying concepts in ALGO-CARE are useful safeguards and it could be adjusted to apply to court officials and algorithm developers.

ALGO-CARE was designed before the implementation of the GDPR and is a working framework for police officers that aims to translate key public law and human rights principles into practical considerations and guidance that can be addressed by public sector bodies.\textsuperscript{386} The original ALGO-CARE framework will be attached as an Appendix. Although this framework was designed for use by police officers, it encompasses principles that could be used by a judiciary in the context of criminal sentencing, and principles, which could arguably fill the gaps and inadequacies of the GDPR and statutory provisions in non-EU countries, including the United States. The framework is an acronym that uses the words ‘Advisory, Lawful, Granularity, Ownership, Challengeable, Accuracy, Responsible and Explainable.’ The framework sets out important questions such as: ‘Does a human officer retain decision-making discretion?’ ‘Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes?’ ‘Is the data categorized to avoid ‘broad-brush’

\textsuperscript{382} Kroll et al (n 22) p. 701
\textsuperscript{383} Ibid.
\textsuperscript{384} Oswald et al (n 143) & see Chapter 1
\textsuperscript{385} See Chapter 1
\textsuperscript{386} Oswald et al (n 143) p. 244
grouping and results, and therefore issues potential bias?’ ‘Does the force need rights to
access, use and amend the source code and data analysed?’ ‘What are post-
implementation oversight and audit mechanisms e.g. to identify any bias?’ ‘Can the
stated accuracy of the algorithm be validated reasonably periodically?’ ‘In what way will
the tool improve the current system?’ ‘Can the percentage of false positives/negatives be
justified?’ ‘Would the operation of the algorithm be considered fair?’ and finally under
Explainable, ‘is appropriate information available about decision-making rule(s) and the
impact that each factor has on the final score or outcome, and is the force able to access
and deploy a data science expert to explain and justify the algorithmic tool (in a similar
way to an expert forensic pathologist)?’

Explainable, is arguably one of the most valuable safeguards in regards to the
context of this research. Explainable is supplemented by another figure, which is
questions and considerations addressing key legal considerations (such as proportionality,
natural justice and procedural fairness), as well as concerns such as intellectual property
ownership and the availability of an ‘expert witness’ to the algorithms functionality in
order to comply with procedural fairness. ALGO-CARE and in particular, the
‘Explainable’ section within the framework, if applied to an algorithm used in criminal
sentencing, would address a significant number of issues that have been mentioned
throughout this research project. Ultimately, however, at present, if ALGO-CARE was
applied to an algorithm such as COMPAS in the context of sentencing, it is unlikely
that it would satisfy the criterion laid out before it, due to the aforementioned human
rights issues that it causes. If ALGO-CARE was applied to COMPAS in Loomis, one
could not deny that the appropriate information about the decision-making rule and the
impact that each factor has on the final score or outcome would not be readily available.
The Court blatantly admitted that they did not understand the decision-making rule of
COMPAS or how the variables were weighted. The Court did not necessarily have access

---

387 Ibid. p. 246 & Appendix figure 3
388 See Appendix figure 3
389 Oswald et al (n 143) p. 246
390 See Chapter 1 & Chapter 2 for information on COMPAS
to a data scientist expert, and when Northpointe offered them the opportunity to oversee an amicus brief, the Court declined.

If the language in ALGO-CARE was altered to address a judiciary and developers creating algorithms for criminal sentencing, rather than a police force, and the framework became part of a statutory regime or national court procedure, it arguably would provide a strong safeguard against due process rights infringements. As it currently stands, it would be difficult to find a criminal justice-based algorithm that could successfully fulfill every condition under the ALGO-CARE framework, thus removing the issue of algorithms used in sentencing impacting due process rights in the first place. It also provides the creators of these tools with guidance in regards to the creation of algorithms specifically for sentencing in the future. The weaknesses to ALGO-CARE as it stands currently are its limited scope and its lack of repercussions, however as an experimental framework, its strengths are certainly something that should be considered by both algorithm creators, data collectors, legislatures, and judiciaries utilizing these tools.

**Practicality & Efficacy of these Safeguards**

While the removal or prohibition of algorithms from the criminal sentencing process would arguably be the most effective safeguard, it is unlikely that the developer’s of the algorithms will be willing to terminate contracts with courts, at least without resistance. U.S. courts will also likely require greater empirical evidence that demonstrates that these tools do present a threat to due process rights; particularly given the precedent they have set through *Loomis*. It may be easier to implement a prohibition on the tools in jurisdictions wherein they are not yet utilized in the sentencing process; however, this would likely be left to policymakers and legislatures to determine whether a prohibition of this technology is warranted. There is a good chance that this safeguard will not be considered an option at all until further research has been conducted.

The safeguard regarding the implementation of mandatory judicial training courses in regards to assessing algorithmic technology is not an unreasonable ask. Given
that there are already courses on the evaluation of evidence, such as forensic evidence, it does not seem unreasonable to assert a need for training regarding the assessment of algorithmic evidence. It merely demonstrates responsibility and adaptation to modern circumstances; however, the main problem with this safeguard would be the ‘mandatory’ element. For example, the aforementioned training courses offered by the Judicial College in the UK are optional and attended based on interest, rather than out of necessity. Thus, there may be hesitation regarding the implementation of something mandatory. One could argue that algorithmic tools are different from evidence such as forensic evidence as they would be applicable in every type of criminal trial, thus the need for mandatory judicial training is enhanced.

Assigning the duty to the same regulator could cut the cost of providing expert testimony and regular audits of algorithmic tools. For example, medical doctors often provide expert testimony in court; however, they are also the same individuals assessing their patient throughout the year. Having an individual who not only understands the algorithmic tool but conducts regular audits on it would be able to provide the court with much more realistic evidence on the tools accuracy and the weighting of particular variables than someone who has never audited the program. The cost of these safeguards may warrant concern, however, the due process obligations would arguably outweigh these concerns. These safeguards are arguably some of the more realistic suggestions, as it does not differ much from the standards required for other evidence such as forensic evidence. One challenge that an auditor or expert witness may face would be whether the developers are willing to release the source code and algorithmic rule or not.

The proprietary nature of these algorithms means that the suggested safeguard of providing open source codes or algorithmic rules may be unrealistic, however, if this safeguard worked in combination with an expert witness, for example, it would be a great success for algorithmic transparency and subsequently due process rights. A contract could be put in place to only provide the proprietary information to expert witnesses in order to maintain secrecy but foster transparency. Realistically, this could result in a court answering the legal question of whether a company’s proprietary secrets outweigh the
need for transparency in algorithmic technology. The GDPR demonstrates a step in the right direction in regards to fostering algorithmic transparency through non-technical language, so its interpretation could assist in the implementation of a safeguard like this.

The final safeguard regarding the design of algorithms with criminal sentencing and due process rights in mind could be highly effective if, for example, the use of immutable characteristics were no longer required as input variables. It may not rid the issue of the right to be sentenced based on accurate information or the right to an individualized sentence, but it could assist with equality challenges under due process. From a practical standpoint, it does not seem unrealistic to expect research to progress in regards to the creation of more accurate and fair algorithmic tools as it currently is a popular topic of debate and discussion globally. However, it is unrealistic to expect this development to be speedy. The implementation of a framework similar to ALGO-CARE would be much more realistic in the near future.\(^{391}\) Moreover, this safeguard would have to work in conjunction with the aforementioned safeguards. As was discussed in Chapter 1, even the best-designed algorithmic tools, may not align with the objectives of public policy.\(^{392}\)

Arguably, most of these safeguards are not dissimilar from practices that already occur in the court process, thus are not overly unrealistic. However, it would be unrealistic to expect their implementation to be speedy and without resistance, particularly regarding releasing source codes and algorithmic rules. The best result for the protection of due process rights would come from a combination of these safeguards, and perhaps in time, algorithmic tools will be treated with more scepticism.

**Conclusion**

The goal of Chapter 4 was to answer the question that surfaced at the beginning of this project: What, if any, amendments could be made to enhance the protection of due

\(^{391}\) See Chapter 3
\(^{392}\) See Chapter 1
process rights from algorithmic technology? The challenges to due process were identified in Chapter 2 and the currently available safeguards were analyzed in Chapter 3. There are clear gaps in the data protection legislation that prevent criminal defendants from utilizing them as grounds to challenge the use of algorithmic tools at sentencing. At the present moment, this project remains an advocate for the removal of algorithms in criminal sentencing based on the identified shortfalls of the available safeguards and risks of their use to a fair trial and due process rights.\textsuperscript{393} However, it is likely that the growing trends towards the more prevalent use of algorithms in various areas of life, will also impact courtrooms and the sentencing process.

The ICCPR has outlined duties on state governments to protect, respect and to take individual steps to support the realization of due process rights. To uphold these duties, states must approach the implementation of algorithmic instruments in criminal sentencing with great caution and provide both statutory and procedural safeguards for defendants. The safeguards in this chapter include a combination of statutory provisions, procedural practices and preventative measures and considerations designed to address the issues of algorithmic opacity, accuracy and the unchallengeable nature of the tools themselves. These safeguards include mandatory judicial training programs on algorithmic evidence, the provision of expert testimony on algorithms, mandatory regular audits of algorithms, and the construction of algorithmic tools with due process in mind.

These safeguards would require cooperation on behalf of agencies administering justice, algorithm developers and experts, and legislatures. The safeguards mentioned above would involve both effort and expense, however, they should not be considered optional given the risk to individual due process and the obligations set by international human rights instruments. Moreover, many of these safeguards including judicial training programs, expert witnesses, and auditing of algorithmic programs are already common practice. There is no reason why defendants should not have their due process rights respected due to the modernity or lack of understanding of algorithms. Safeguards should not be optional in the quest to protect and respect due process rights.

\textsuperscript{393} See Chapter 2 & Chapter 3
Conclusion

The purpose of this project was to explore the use of algorithms in criminal sentencing and analyze their impact and consequences on due process rights. This project employed an interdisciplinary approach, using socio-legal analysis and applied a human rights lens to the literature gathered. This project was divided up into four sections to address the four following questions: (1) how are algorithms and automated data processing systems utilized in the context of criminal sentencing? (2) How does the use of algorithms in criminal sentencing impact due process rights? (3) What safeguards are available to those subject to automated data processing at criminal sentencing and how effective are they? (4) What, if any, amendments could be made to enhance the protection of due process rights from algorithmic technology?

The first Chapter provided a review of the current literature available on the use of algorithms in criminal sentencing and the criminal justice system more generally. This chapter analyzed the empirical studies conducted on algorithmic tools in this context and also provided the theoretical and practical justifications for utilizing these tools in sentencing. Moreover, this chapter included a detailed analysis of various case law examples wherein the use of an algorithmic tool at sentencing was challenged in court. At present, there have currently been no case examples wherein the court has ruled that the use of algorithmic tools at sentencing should be prohibited on due process grounds. In fact, there has yet to be a case wherein the court has found any human rights violation whatsoever in using these tools at sentencing. Finally, this chapter analyzed the literature on the structural issues of algorithmic tools. Through this analysis of the literature, it was identified that opacity, immutable characteristics used as input variables and varied rates of statistical accuracy were serious threats to due process rights. Moreover, the lack of general understanding of algorithms and their technological rules on behalf of the judiciary and defendant’s rules was also found to be problematic for challenging the tools.
Chapter 2 addressed the question: How does the use of algorithms in criminal sentencing impact due process rights? This chapter provided a brief background on the development of due process rights and concluded that despite the differences in formal conceptions of due process, the rights included in the notion of due process are some of the most agreed-upon rights internationally. This chapter employed various international and national human rights instruments, along with case law, to assess the impact of algorithms in sentencing on due process rights. The application of various conceptions of due process rather than just the rights provided under the United States Constitution was one of the major contributions to research in this area as the application of other instruments such as the ECHR and its case law is generally absent from the literature.

The first right wherein the impact of algorithms was analyzed was the right to review and verify one’s sentence. This analysis concluded that the lack of transparency and understanding of algorithmic tools left defendants without the ability to review and verify their sentences. Next, the implications of the right to an individualised sentence were analyzed and it was determined that inferring individual tendencies from group data may undermine this right as it is presented in case law. Then, this project moved to assess the impact of utilizing particular variables as input data for the algorithms such as race, gender, socioeconomic status, education and employment. It was found that the use of each variable, and in some cases the removal of the variable, would result in biased outcomes, which would impact one’s right to a fair trial and equality under the law. Overall, it was found that the use of algorithmic tools in criminal sentencing is a threat to due process and may undermine individual rights including, but not limited to, the right to review and verify sentencing, the right to an individualized sentence, and the right to a fair trial and equality under the law.

Having considered the implications of algorithmic tools on due process rights, the purpose of Chapter 3 was to answer the question: What if any safeguards are available to those who are subject to algorithmic tools at sentencing, and how effective are these safeguards? As data protection instruments protect various human rights, the recently implemented GDPR, its Directive and the UK’s DPA 2018 were analysed to see if they
could be of any use to defendants subject to algorithmic tools at trial. Various provision’s relevant to the subject matter were analysed and applied to both hypothetical cases, and the *Loomis* case. The findings of this analysis were that this legislation is of little value to defendants subject to algorithmic tools at trial. While these instruments are a step in the right direction in regards to the protection of human rights generally, they pay little attention to due process rights. This section also included a brief analysis of the United State’s approach to data protection and human rights and concluded that the U.S. sectoral approach and focus on consumer relations does not provide defendants with any form of safeguard from algorithmic harms at sentencing. Overall, it was concluded that the currently available safeguards are ineffective and do not adequately protect due process rights from algorithmic harms in criminal sentencing.

The final chapter of this project combined the knowledge gathered from the three previous chapters of this project and utilized it to develop recommendations of safeguards for protecting the due process rights of defendants subject to algorithms at trial. While this project advocates for the removal of algorithms from criminal sentencing in the current circumstances, the author acknowledges that the growing trend of algorithms being utilized in new areas of life will likely spread to courtrooms, beyond the United States. Thus, the safeguard recommendations and suggestions provided should be considered prior to the implementation of these tools in courtrooms. The safeguards that are suggested were based on the due process challenges identified in Chapter 1 and 2, and also the gaps that were identified in the currently available data protection legislation. These safeguards included the implementation of mandatory judicial training on assessing algorithmic evidence, the implementation of expert testimony for algorithmic tools, regular auditing of algorithms, and designing algorithms with due process and criminal sentencing at the forefront.

If the answers to the research questions contained within this project could be summed up in one sentence it would be: Algorithms in criminal sentencing are a threat to due process rights and as it currently stands, there are no adequate safeguards for the protection of these rights. This project contributed to the literature as it provided an
analysis of the issue of the impact of algorithms in due process rights in sentencing using human rights instruments other than the United States Constitution. The absence of the literature addressing algorithms in sentencing outside of the United States context was challenging, however, it provided space for unique application of human rights and data protection instruments to the identified due process challenges. This unique perspective also assisted in developing suggestions and recommendations for safeguarding due process rights in the future.

The Research Outlook

As mentioned in the concluding remarks, there is a large gap in the literature wherein the due process implications of using algorithmic tools at criminal sentencing have been analysed in general, but in particular, outside of the context of the United States Constitution. While this project provided some original analysis and application of international human rights instruments to the identified due process challenges, this issue should be researched further as the existing scholarship has not been attentive enough. Further, the proposed safeguards and recommendations laid out in Chapter 4 are based on the identified due process challenges and gaps in the currently available safeguards, but their efficacy would have to be tested post-implementation. As there has yet to be a national Supreme Court ruling in any jurisdiction on the violation of due process rights from the use of an algorithm at criminal sentencing, the research and outcome of that ruling will be much anticipated, should it ever come.
Bibliography

Secondary Sources: Academic Articles & Journal Articles


Bagaric, M., and Wolf, G., ‘Sentencing by computer: enhancing sentencing transparency and predictability, and (possibly) bridging the gap between sentencing knowledge and practice,’ 2017 George Mason Law Review 25(4) 1-50


Carr, N. and Maguire, N., ‘Pre-sentence reports and Individualised Justice: Consistency, Temporality and Contingency,’ 2017 *Irish Probation Journal* 14 52-71


Edwards, L. and Veale, M., ‘Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy you are Looking For,’ 2017 Duke Law and Technology Review 16(1) 1-65

Eijk, G.V., ‘Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality,’ 2017 Punishment & Society 19(4) 463-481


Gerards, J., ‘The fundamental rights challenges of algorithms,’ 2019 Netherlands Quarterly of Human Rights 37(3) 305-209


Green, B., “”Fair” Risk Assessments: A Precarious Approach for Criminal Justice Reform,’ 2018 Harvard University In 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018)

Hannah-Moffat, K., ‘Criminogenic Needs and the Transformative Risk Subject,’ 2005 *Punishment & Society* 7(1) 29-51


Hannah-Moffat, K., ‘Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates,’ 2018 *Theoretical Criminology* 23(4) 453-470


Hart, S.D., Michie, C. and Cooke, D.J., ‘Precision of actuarial instruments: evaluating the ‘margins of error’ of group v. individual predictions of violence,’ 2007 *British Journal of Psychiatry* 190(49) 60-65

Herrschafft, B.A., ‘Evaluating the reliability and validity of the COMPAS tool: implications for community corrections policy,’ 2014 *RUcore: Rutgers University Community Repository*


Kort F., ‘Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the Right to Counsel Cases,’ 1975 American Political Science Review 51(1) 1-12


Lauderdale, B.E. and Clarke, T.S., ‘The Supreme Court’s Many Median Justices,’ 2012 American Political Science Review 106(4) 847-866


Monahan, J. and Skeem, J.L., ‘Current directions in violence risk assessment,’ 2011 *Current Directions in Psychological Science* 20(1) 38-42


O’Connell, R., ‘Cinderella comes to the Ball: Article 14 and the right to non-discrimination in the ECHR,’ 2009 *Legal Studies* 29(2) 211-229


Radacic, I., Gender Equality Jurisprudence of the European Court of Human Rights,’ 2008 *European Journal of International Law* 19(4) 841-857


Schomburn, W., ‘The Role of International Criminal Tribunals in Promoting Respect for Fair Trial Rights,’ 2009 *Northwestern Journal of International Human Rights* 8(1) 1-29


Tata, C., ‘Ritual Individualization’: Creative Genius at Sentencing, Mitigation, and Conviction,’ 2019 *Journal of Law and Society* 46(1) 112-140


**Secondary Sources: Books & Book Chapters**


Davis, G.R.C., ‘Magna Carta’ Revised Ed. (British Library Publications 1989)


Skeem, J., and Mulvey, E.P., ‘Monitoring the violence potential of mentally disordered offenders being treated in the community,’ in Care of the Mentally Disordered Offender in the Community (Oxford University Press 2002)


**Secondary Sources: Reports, Guidelines and General Comments**

American Bar Association, ‘State Policy Implementation Project,’ *Criminal Justice Section:*

<https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_civilcitations.authcheckdam.pdf> accessed: January 13, 2019


Borgesius, F.Z., ‘Discrimination, artificial intelligence, and algorithmic decision-making,’ 2019 *Council of Europe*

Committee of Experts on Internet Intermediaries (MSI-NET), ‘Algorithms and Human Rights: study on the human rights dimensions of automated data processing techniques and possible regulatory impacts,’ 2017 *Council of Europe*
<https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a>

Dietrich, W., Mendoza, C., and Brennan, T., COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity,’ 2016 *Northpointe Report*


Human Rights Committee, ‘General Comment No. 32: Article 14: Right to equality before courts and tribunals and to a fair trial,’ 2007 *UN Doc CCPR/C/GC/32* 90th Session
James, N., ‘Risk and Needs Assessment in the Federal Prison System,’ 2018


Urwin, S., ‘Written evidence for the inquiry into Algorithms in Decision-making,’ 2018 Head of Criminal Justice Durham Constabulary

Secondary Sources: Websites and Press Releases


Courts and Tribunals Judiciary, ‘The Judicial College and training,’ 2019
<https://www.judiciary.uk/about-the-judiciary/international/training-for-overseas-judges/> Accessed August 23, 2019


European Court of Human Rights, ‘Life sentencing in Russia is not discriminatory,’ 2017 Press Release issued by the Registrar of the Court ECHR 003 < https://hudoc.echr.coe.int/app/conversion/pdf/?library=ECHR&id=003-5607491-7086661&filename=Grand%20Chamber%20judgment%20Khamtokhu%20and%20Aksenchik%20v.%20Russia%20-%20life%20sentencing%20in%20Russia%20is%20not%20discriminatory.pdf> accessed August 1, 2019


<https://www.ohchr.org/EN/ProfessionalInterest/Pages/InternationalLaw.aspx>
accessed: August 1, 2019

United Nations Office of the High Commissioner, ‘Death penalty disproportionately affects the poor, UN rights experts warn,’ 2017 <

Primary Sources: Case Law

United Kingdom:
   Essop v Home Office (UK Border Agency) (2017) UKSC 27, SC
   R v Commissioner of Police of the Metropolis and another (2011) UKSC 9

United States:
   Chambers v. Florida (1940) 309 U.S. 227
   Griffin v Illinois [1956] 351 U.S. 12
   State of Iowa v. Sean David Gordon (2018) CA Iowa
   State of Wisconsin v. Loomis (2016) 881 N.W.2d 749 (Wis. 2016)
   United States v Maples (1974) 501 F.2d 985
   United States v Virginia (1996) 518 U.S. 515
Washington v. Davis (1976) 426 U.S. 229
Williams v Currie [2000] M.D.N.C
Williams v. New York (1949) 337 U.S. 241

Canada:
R v. Gladue (1991) 1 S.C.R 688

Australia:
Director of Public Prosecutions for Western Australia v Mangolamara (2007) 169 A Crim R [2007] WASC 71
Elias v. The Queen (2013) 248 CLR 483

European Court of Human Rights Case Law:
Abdulaziz, Cabeles and Balkandali v. The United Kingdom application no. 9214/80 (ECtHR 28 May 1985)
Ecis v. Latvia application no. 12879/09 (ECtHR 10 January 2019)
Khamtokhu and Asenchik v Russia application no. 60367/08 (ECtHR 24 January 2017)
Konstantin Markin v. Russia (Grand Chamber judgment of 22 March 2012)

European Union Case Law:
Case-54/07, Centrum voor gelijkheid van khasen en voor racismebestrijding v. Firma Feryn NV 2008

Primary Sources: Legislation

UK legislation
Data Protection Act 2018 c. 12
Equality Act 2010 c. 15
Magna Carta 1219 c.9

United States Legislation

California Consumer Privacy Act 2018 (Assembly Bill 375)

United States Constitution 1787

Canada Legislation


Corrections and Conditional Release Act (S.C. 1992, c.20)

Australia Legislation

Dangerous Sexual Offenders Act 2006 (WA)

European Union Legislation


International Human Rights Instruments

Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights, as amended) (ECHR)

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR)

Universal Declaration of Human Rights (adopted 10 December 1948 UNGA Res 217 A(III) (UDHR)
Appendix 1: Preparatory Case Note

*State of Wisconsin v. Eric L. Loomis [2016] 881 N.W. 2d 749*

*Algorithms—Due Process Rights—Criminal Sentencing—Wisconsin State Law*

**Facts**

In early 2013, the State of Wisconsin contended that Eric L. Loomis was the driver in a drive-by shooting. Loomis was charged with five counts, all as a repeater of: (1) First-degree recklessly endangering safety; (2) Attempting to flee or elude a traffic officer; (3) Operating a motor vehicle without the owner’s consent; (4) Possession of a firearm by a felon; and (5) Possession of a short-barreled shotgun or rifle. Loomis denied involvement in the drive-by shooting and waived his right to a trial by entering a guilty plea to only two lesser offences including attempting to flee a traffic officer and operating a motor vehicle without the owner’s consent.

The circuit court accepted Loomis’s plea and ordered a presentence investigation (PSI). The PSI included a COMPAS risk assessment. COMPAS assessments estimate the risk of recidivism based on information gathered from the defendant’s criminal file and an interview with the defendant. Given that the methodology behind COMPAS is a trade secret, only the estimates of recidivism risk are reported to the court. In addition to the COMPAS assessment, the circuit court considered the read-in charges at sentencing and assumed that the factual bases for the read-in charges were true, noting the serious and aggravating nature of shooting related charges.

At the sentencing hearing, the circuit court referred to the COMPAS assessment in its sentencing determination and, based in part on the COMPAS assessment, sentenced Loomis to six years imprisonment and five years of extended supervision. The circuit court

---

394 (2016) 881 N.W. 2d 749
395 Ibid. 7
396 Ibid
397 Ibid.
court also denied Loomis’s motion for post-conviction relief requesting a resentencing hearing.

Loomis appealed to the Wisconsin Supreme Court on the grounds that the circuit court’s use of the COMPAS risk assessment at sentencing violates a defendant’s right to due process for three reasons: (1) it violates a defendant’s right to be sentenced based upon accurate information, in part because the proprietary nature of COMPAS prevents him from assessing its accuracy; (2) it violates a defendant’s right to an individualized sentence; and (3) it improperly uses gendered assessments in sentencing. Additionally Loomis contended that the circuit court erroneously exercised its discretion by assuming that the factual bases for the read-in charges were true.

Issues

Whether it is a violation of a defendant’s constitutional right to due process for a trial court to rely in part on the risk assessment results provided by a proprietary risk assessment instrument such as the Correctional Offender Management Profiling for Alternative Sanctions at sentencing.

Outcome

The Wisconsin Supreme Court affirmed the circuit court’s sentence concluding that if used properly, observing specific limitations and cautions set forth therein, a circuit court’s consideration of a COMPAS risk assessment tool at sentencing does not violate defendants right to due process. Moreover, the circuit court did not erroneously exercise its discretion through the reading-in of charges because ‘it employed recognized legal standards.’ The Supreme Court denied Loomis’s motion for post-conviction relief.

---

398 Ibid. 13, 14
399 Ibid. 5
Judgment

Justice Ann Walsh Bradley rejected Loomis’s due process arguments. First, the argument put forward by Loomis that analogized the use of the COMPAS algorithm to cases such as *Gardner* and *Skaff* and noted that the defendant could not challenge or review how the algorithm calculated risk was deemed ‘imperfect.’ Northpointe is the developer of the COMPAS tool. Justice Bradley referenced Northpointe’s 2015 Practitioner’s Guide to COMPAS noting that the risk scores are based largely on static information such as the criminal history of the defendant, with limited use of dynamic variables such as the defendant’s criminal associates or substance abuse. Further the COMPAS report included a list of 21 questions regarding the static factors, which were accessible by Loomis in his PSI report. Justice Bradley argued that Loomis had access to this information and had the opportunity to challenge his risk scores by arguing that other factors or information demonstrated its inaccuracy. Justice Bradley did, however, state that regardless of whether *Gardner* and *Skaff* are analogous to this case at hand, Loomis correctly asserted that defendants have the right to be sentenced based on accurate information as per *Travis*. This is where various studies are raised regarding the accuracy of the COMPAS algorithm. Justice Bradley noted that the ‘state of Wisconsin has not yet completed a statistical validation study of COMPAS for a Wisconsin population;’ and cited a study by the California Department of Corrections (CDRC) that found that ‘there is little evidence that this COMPAS is actually assessing criminogenic needs and that it could not be recommended that CDRC uses COMPAS for individual sentencing.’ Additional concerns are raised about risk assessment tools and that they may disproportionately classify minority offenders as higher risk, often due to factors that may be outside of their control such as familial background and education.

---

400 Ibid 22  
401 Ibid. 22  
402 Ibid. 23  
403 347 Wis. 2d 142, 17  
404 *State v. Loomis* (n 394)  
In response to the studies that were critical of risk assessment tools, Justice Bradley set out specific limitations and cautions for sentencing courts utilizing tools such as COMPAS. These included: (1) informing the court of the proprietary nature of COMPAS, which prevents disclosure of information relating to how factors are weighed or how risk scores are assessed. (2) Informing the court that the risk assessment is based on a national sample, but no cross validation study for a Wisconsin population has occurred. (3) Informing the court that there are various studies that indicate that COMPAS disproportionately classifies minority offenders as having a higher risk of recidivism. (4) Risk assessment tools must constantly be monitored and re-normed for accuracy due to changing populations and subpopulations; and (5) COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole.\textsuperscript{406} Justice Bradley argued that providing courts with this information in written form would enable them to better assess the accuracy of the assessment and the appropriate weight to be given to the risk score.\textsuperscript{407}

In response to Loomis’s second argument that the court’s consideration of a COMPAS assessment violated his right to an individualized sentence, Justice Bradley emphasized the importance of individualized sentencing and admitted that COMPAS only provides aggregate data on recidivism risk for groups similar to the offender, but she also explained that the assessment would not be the sole basis for a sentencing decision. Sentencing that includes a COMPAS assessment would be sufficiently individualized because courts have the information to disagree with the assessment, and that the Court ‘expects’ that circuit courts will exercise discretion when assessing risk scores with respect to each individual defendant.\textsuperscript{408} Further, Justice Bradley disagreed with Loomis because ‘consideration of a COMPAS risk assessment at sentencing along with other

\textsuperscript{406} State v. Loomis (n 394) 28
\textsuperscript{407} Ibid.
\textsuperscript{408} State v. Loomis (n 394) 30
supporting factors is helpful in providing the sentencing court with as much information as possible in order to arrive at an individualized sentence.409

A case that was referenced often throughout this decision was Gallion. Gallion was a case involving a defendant who claimed his due process rights had been breached as the circuit court erroneously exercised its discretion by failing to provide an adequate explanation for the sentence and relying too heavily on his character.410 New sentencing regulations in Wisconsin had abolished parole boards. Justice Bradley cited Gallion arguing that the ‘circuit courts have an enhanced need for more complete information upfront, at the time of sentencing,’411 thus justifying the need for greater background and character information on the defendant due to the parole boards abolishment. In Loomis, the judges held the view that COMPAS assisted in providing more information on the defendant.

Finally, in response to Loomis’s third argument regarding the use of gender as an variable in COMPAS risk assessment violating the right to due process, Justice Bradley noted that the use of this variable promotes accuracy that ultimately works to the benefit of the justice system. Justice Bradley further argued that Loomis failed to meet his burden of proof demonstrating that the sentencing court actually relied on gender as a factor in sentencing as per Harris.412 Thus the court concluded that the use of the COMPAS risk assessment tool did not violate Loomis’s right to due process, but set forth a number of permissible uses of the COMPAS tool and then limitations and cautions. The permissible uses of COMPAS in a sentencing court included (1) diverting low-risk prison-bound offenders to a non-prison alternative; (2) assessing whether an offender can be supervised safely and effectively in the community; and (3) imposing terms and conditions of probation, supervision, and responses to violations, however, the risk score cannot be determinative.413 Justice Bradley stated that the COMPAS risk assessment

409 Ibid.
410 State v Gallion (2004) Wis. 42
411 State v. Loomis (n 394) 30
412 326 Wis. 2d 685
413 State v. Loomis (n 394) 36
might be used to ‘enhance a judge’s evaluation, weighing, and application of the other sentencing evidence in the formulation of an individualized sentencing program appropriate for each defendant.’ 414

The Supreme Court judges set forth more limitations on the use of COMPAS as well as the requirement to provide circuit courts with the written advisement if COMPAS is used in the PSI. These limitations included remembering that risk scores are not intended to determine the severity of the sentence or whether an offender is incarcerated; also that risk scores are not to be used as the determinative factor in deciding whether the offender can be supervised safely and effectively in the community. 415 Additionally, Justice Bradley noted that a COMPAS risk assessment was not designed to address all of the goals of sentencing which include deterrence; rehabilitation; retribution; and segregation as per State v Dowdy. 416 Justice Bradley stated that because of these disparate goals, using a risk assessment tool to determine the length and severity of a sentence would be a poor fit. 417 Further, the circuit court is required to explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. Justice Bradley emphasized that these cautions and limitations are suitable at the present time, however, they are adjustable and should be updated on a regular basis.

The Supreme Court then applied the aforementioned cautions and limitations to the case at hand and found that, although the circuit court was unaware of those newly created cautions, the circuit court had referenced the risk assessment at sentencing and essentially gave it little or no weight, therefore, not violating Loomis’s right to due process. The Supreme Court also dismissed Loomis’s appeal on the grounds of the read-in charges as they found that the circuit court had weighed the facts, assessed the credibility and recognized legal standards for read-in offenses, and found that the circuit

414 Ibid. 38
415 Ibid.
416 [2012] 338 Wis.2d 565
417 State v. Loomis (n 394) 39
courts consideration of read-in charges was not an erroneous exercise of discretion. Accordingly, the Supreme Court affirmed the order of the circuit court.

Justice Drake Roggensack concurred, emphasizing that consideration of COMPAS in sentencing is permissible, however, reliance on COMPAS for the sentence imposed is not permissible and would violate due process protection. Justice Shirley S. Abrahamson also concurred; however, she emphasized the court’s lack of understanding of COMPAS, and it being a significant problem. Northpointe sought to file an amicus brief in the case at hand to discuss the history, accuracy and efficacy of COMPAS throughout the criminal justice system, however, the court denied Northpointe’s motion to file an amicus brief, and Justice Abrahamson called this a ‘mistake.’ In light of the criticisms that these data driven tools have drawn from officials and scholars, Justice Abrahamson argued that ‘a more extensive record from sentencing courts on the strengths, weaknesses and relevance to the individualized sentencing process of these evidenced based tools is necessary to provide the defendant, State and public with a transparent and comprehensible explanation for the sentencing court’s decision.’ There were no dissenting opinions in the case at hand.

Comment

On October 5, 2016 a petition for a writ of certiorari was filed to the United States Supreme Court to overturn the ruling in the State of Wisconsin v. Eric L. Loomis [2016], however, the United States Supreme Court denied the writ of certiorari and declined to hear the case on June 26, 2017. ‘Loomis is a landmark case on the legality of using risk assessment tools in a sentencing court and unsurprisingly, it quickly hit the headlines and

418 Ibid 47
419 Ibid. 4 (in Roggensack J.concurring judgment)
420 Ibid. 2 (in Abrahamson J. concurring (judgement)
421 Ibid. 5
stimulated enormous debate among scholars, practitioners and policy makers.\(^4\) The Loomis case arguably could be said to be an attempt to curb the enthusiasm surrounding the use of risk assessment tools in criminal sentencing in the United States. Justice Bradley demonstrated this enthusiasm by citing the Indiana Supreme Court case of _Malenchik v. State_ wherein the court found the assessment tools assist courts in weighing all the sentencing factors:

‘Such assessment instruments enable a sentencing judge to more effectively evaluate and weigh several express statutory sentencing considerations such as criminal history, the likelihood of affirmative response to probation or short term imprisonment, and the character and attitudes indicating that a defendant is unlikely to commit another crime.’\(^4\)

Moreover, the question of whether COMPAS could be used at sentencing was also addressed previously by the Wisconsin Court of Appeal in _the State of Wisconsin v. Samsa_ wherein the court approved of the circuit court’s consideration of a COMPAS assessment in sentencing,\(^4\) however, it was not presented with the same due process implications as _Loomis_. Numerous states use risk assessment tools at sentencing and some even require judges to consider them when making decisions. The Virginia Criminal Sentencing Commission and Utah Sentencing Commission have developed and applied risk assessment tools that are used to assess non-violent offenders bound for incarceration.\(^4\) Further, the American Bar Association has urged states to adopt risk

\footnotesize{\textsuperscript{4} Han-Wei Liu, Ching-Fu Lin & Yu Jie Chen, ‘Beyond _State v Loomis_: artificial intelligence, government algorithmization and accountability,’ 2019 _International Journal of Law and Information Technology_ 27(2) 122-141 p. 8
\textsuperscript{4\textsuperscript{2}} _Malenchik v. State_ 928 N.E. 564, 574
\textsuperscript{4\textsuperscript{2}5} [2015] WI App 6, 359 Wis. 2d. 580, 859 N.W.2d 149
assessment tools in an effort to reduce recidivism and increase public safety.\textsuperscript{427} These are just a few examples of how courts in the United States are enthusiastically adopting these tools in sentencing.

The Wisconsin Supreme Court brought attention to the lower courts of the opacity of these algorithms and provided a list of limitations and cautions that are open to change; however, one could argue that these cautions are insufficient to curb the apparent enthusiasm for the use of risk assessment tools in sentencing. Justice Bradley focused heavily on the circuit court’s ability to dismiss risk assessment scores, or not to be heavily influenced by them. However, it may be unrealistic to expect a trial judge to disregard risk scores attached to PSI’s given the numerous studies demonstrating how humans are submissive to computer-generated numbers and results that may further frame and condition the view of judges.\textsuperscript{428} Dr Starr remarked that it is hard to think of a situation where a ‘high risk’ label will not result in a longer sentence.\textsuperscript{429}

‘Mere written warnings do not seem to be able to adequately inform judges as effective gatekeepers, especially when they may not be sufficiently equipped with expertise as to understand the workings of such tools.’\textsuperscript{430} Justice Abrahamson even commented on the deep lack of understanding of the workings of tools such as COMPAS, however, the Court, insensitive to this reality itself, accepted the claim of the circuit court that it would have imposed the exact same sentence, regardless of the COMPAS risk scores.\textsuperscript{431} Rather than placing any restraint on the use of COMPAS, the court chose to trust members of the judiciary to not be swayed by a COMPAS output.

The Court acknowledged the lack of transparency due to the proprietary nature but then accepted the output of the algorithm based on the fact that the data and information

\textsuperscript{427} American Bar Association, ‘State Policy Implementation,’\textit{ Criminal Justice Section} retrieved: https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_civilcitations.authcheckdam.pdf
\textsuperscript{428} Liu et al (n 423) 9
\textsuperscript{430} Liu et al (n 423) 10
\textsuperscript{431} Ibid.
provided by Loomis was public and accessible to both the Court and the defendant himself. However, no explanation of the weighing of these variables or how the input affects the output was ever provided to the Court or to Loomis. There was no way for the Court or Loomis to understand how COMPAS weighed each variable or understand why Northpoint chose to assign individual variables their weight. Thus, only the output and input of COMPAS could be analysed. The main problem with this is that it prevents Loomis from being able to meaningfully challenge the algorithm, like he could challenge an expert witness for example. Further, in pointing to Northpointe’s practitioner guide as a defendant’s tool for assessing the accuracy of the score, the court blatantly ignored the fact that Northpointe is a for-profit company with a $1,765,334 contract at stake in Wisconsin’s use of their services.432 ‘With that sum of money tied to the use of COMPAS, Northpointe is thereby a biased party and cannot be relied upon to determine the accuracy of the risk assessment score.’433

The Court’s lack of knowledge surrounding processing and the computation phase of algorithmic tools demonstrates the problem of COMPAS and other risk assessment tools lack of transparency, which is described by Frank Pasquale as the ‘black box’.434 ‘The problem of the black box refers to the secrecy of the algorithmic process, which thwarts meaningful scrutiny of algorithms and automated decision-making process systems that have an immense impact on society.’435 The Court failed to address the ‘black box’ challenge and its legal ramifications and arguably the Loomis decision could impede the role of the judiciary as the final gatekeeper to protect individual’s due process rights and to provide effective remedies.436 Perhaps, if Northpointe had have been allowed to file their amicus brief, it would have provided an opportunity to provide greater insight into the ‘black box’ of the COMPAS algorithm. Although large companies producing these risk assessment tools may object to complete public disclosure of their algorithms, an

433 Ibid. p. 92
435 Liu et al (n 423) p. 13
436 Liu et al (n 423) p. 14
alternative could be to disclose only to interested parties or an expert committee. Given the importance of the protection of due process rights and public interest, secrecy for profit should be reasonably confined.\textsuperscript{437} Furthermore, having complete access to the source code would allow for defence attorneys to employ their own experts and evaluate the scores, much like those conducted on the source codes of DNA algorithms and breathalyzers, which can aid in finding errors and ultimately lead to more reliable and accurate technology.\textsuperscript{438}

Another consideration regarding the ‘black box’ challenge would be the way in which the Wisconsin Court interpreted ‘accurate information.’ The Court held that Loomis had the right to be sentenced based on ‘accurate information,’ but they interpreted ‘accurate information’ to mean that the risk assessment must be statistically accurate.\textsuperscript{439} Arguably, however, the Wisconsin Court should not have been focusing on statistical accuracy when assessing the accuracy of sentencing information. According to Lightbourne, the validity measurements that these tools rely on, known as the area under the curve, relies on the ratio of false negatives to false positives. The industry accepted standard is $ROC = .70$, meaning a defendant is correctly classified only 70 percent of the time.\textsuperscript{440} Moreover, there is a 30 percent chance that a lower risk individual would be ranked higher than an actual high-risk individual and vice versa. As the United States criminal justice system is premised on the theory that ‘it is far worse to convict an innocent man, than to let a guilty man go free,’\textsuperscript{441} basing sentencing decisions on these rates is unnerving.\textsuperscript{442}

The way in which the Wisconsin Supreme Court handled the argument of a risk assessment tool violating Loomis’s right to an individualized sentence was substandard compared to other Courts. Remarking that COMPAS would not be the sole factor in the sentencing decision does not resolve the issue that a defendant will automatically be

\textsuperscript{437} Liu et al (n 423)
\textsuperscript{438} Freeman (n 432)
\textsuperscript{439} State v. Loomis (n 394) Justice Bradley quotes various studies regarding the statistical efficiency and accuracy of COMPAS
\textsuperscript{441} re Winship [1970] 397 U.S. 358
\textsuperscript{442} Lightbourne (n 440)
compared to a large subset of the population, rather than be judged on his own merits and social circumstances. The Supreme Court of Western Australia contrasted Loomis in *Director of Public Prosecutions for Western Australia v Mangalomara* where they highlighted the fact that risk assessment tools ‘were not devised for and do not necessarily take account of the social circumstances of indigenous Australians in remote communities,’ and concluded that they had ‘grave reservations as to whether a person of the respondent’s background can be easily fitted within the categories of appraisal presently allowed for by the risk assessment tools.’\(^{443}\) The Supreme Court of Western Australia also noted that ‘the research data and methods underlying the assessment tools are assumed to be correct but this has not been established by the evidence.’\(^{444}\) This is a demonstration of the importance that the Western Australian Court put on an individualized sentence and its regard for social circumstances and unique demographics. Given the massive overrepresentation of African Americans in United States prisons, and the fact that various studies have concluded that COMPAS skewed towards labeling black defendants as high risk and white defendants as low risk despite not using ethnicity as a variable,\(^{445}\) one would assume, or at least be hopeful in thinking that the Wisconsin Supreme Court would prioritize an individualized sentence as the Australian Court has done, over the mandatory use of a risk assessment tool like COMPAS. The *Loomis* case failed to address the individualized sentencing issue and while Justice Bradley highlighted the aforementioned studies regarding skewed results based on race and other variables out of the defendants control, the Court deemed risk assessment tools constitutional.

**The Impact of Loomis**

The *Loomis* case was cited in the *State of Iowa v. Gordon*, where the defendant pleaded guilty to statutory rape. The defendant appealed his prison sentence on the grounds that the district court improperly decided he should be incarcerated based, in

\(^{443}\) *Director of Public Prosecutions for Western Australia v Mangalomara* [2007] 169 A Crim R [2007] WASC 71 [165]

\(^{444}\) Ibid.

\(^{445}\) Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, “How we analyzed the COMPAS recidivism algorithm,” 2016 *ProPublica*
part, on his risk level scores derived from two sex-offender risk-assessment tools included in his PSI report. Gordon used *Loomis* in his argument citing that although Iowa court’s had not yet addressed the proper use of risk assessment tools in sentencing, the decision of the district court ran afoul of the cautions and limitations set forth in *Loomis*. The Iowa Court of Appeal reversed and remanded the case for resentencing, as they found no statutory authority for the use of the risk assessment scores in sentencing. The Iowa Court’s majority opinion was directly contrary to *Loomis* as they found that the district court abused its discretion in considering the risk assessment scores as an ‘aggravating factor’ in deciding the defendant should be incarcerated rather than released into the community with supervision. The Iowa Court specifically stated its opposition to *Loomis*, as it cannot be interpreted to prohibit the sentencing court’s use of risk assessment information in ordering incarceration because Loomis affirmed the defendant’s sentence of incarceration. The Iowa Court did not reverse the finding on constitutional/due process grounds, rather, they focused on the fact that these risk assessment instruments were not designed for use in sentencing and that nothing in their record indicates the existence of validation studies for these risk assessment tools or any cross validation for an Iowa population of offenders. This Court appeared to have a greater regard for individualized sentencing and accuracy of information than the *Loomis* court did, despite deciding on non-constitutional grounds, and it will be interesting to see if future courts arrive at decisions directly contrary to that of *Loomis*.

To its merit, the Court in *Loomis* did acknowledge the issues regarding the validity of COMPAS and provided the list of cautions and limitations, however, its solution does not address the fact that defendants do not have the resources to effectively challenge, or analyze the accuracy of COMPAS to the extent that due process would require. Further, the Court failed to allow for any form of insight into the technological ‘black box’ of the COMPAS algorithm by dismissing an amicus brief from Northpointe. Ultimately, the

---

446 *State of Iowa v. Sean David Gordon* [2018] CA Iowa
447 Ibid.
448 Ibid.
449 Ibid.
450 Ibid.
Wisconsin Supreme Court failed to protect due process rights to the best of their ability and although it may be a landmark case that somewhat curbs the enthusiasm of the use of risk assessment tools in the courtroom, it is hard to say whether the warnings of the court will fall on deaf ears.

The *Gordon* case arrived at a decision contrary to *Loomis* but did not go far enough as to deem the use of risk assessment tools in criminal sentencing as unconstitutional. Perhaps it should be made a requirement that companies such as Northpointe be legally required to provide an in-depth explanation as to the weighting of their variables and the relationship between the input and output of algorithms like COMPAS if they are being used during sentencing or perhaps they should be removed from trial completely. ‘To acknowledge these risks might mean removing algorithms from the sentencing process all together until extensive training takes place and more safety measures are arranged, but losing the efficiency of the algorithms is a low cost to bear when due process rights are at stake.’\(^{451}\) Should Courts still wish to reap the apparent benefits of convenience or additional information from these algorithms, safeguards must be put in place to protect defendant’s due process rights. Even the choice to opt-out of being subject to these algorithmic tools may be of benefit. The lack of empirical studies on the effectiveness and accuracy of these tools in sentencing means that it is too risky to follow the current trend of implementing these tools in sentencing.

\(^{451}\) Freeman (n 432)
Bibliography

Primary Sources: United States Case Law

Malenchik v. State [2010] 928 N.E. 564, 574
re Winship [1970] 397 U.S. 358
State v. Dowdy [2012] 338 Wis.2d 565
State of Wisconsin v. Gallion [2004] 678 N.W. 2d 197
State of Wisconsin v. Harris [2010] 786 N.W. 2d 409
State v. Skaff [1989] 447 N.W. 2d 84
State v. Travis [2013] 832 N.W. 2d 491

Primary Sources: Australian Case Law

Director of Public Prosecutions for Western Australia v Mangolamara [2007] 169 A Crim R [2007] WASC 71 [165]

Secondary Sources:

American Bar Association, ‘State Policy Implementation Project,’ Criminal Justice Section:
<https://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_civilcitations.authcheckdam.pdf> accessed: October 13, 2018

Larson, J., Mattu, S. and Angwin, J., “How we analyzed the COMPAS recidivism algorithm,” 2016 ProPublica


The General Data Protection Regulation (GDPR) 2016/679 is a regulation in European Union law on data protection and privacy for all individuals within the European Union and the European Economic Area. It also addresses the export of personal data outside the EU and EEA areas. It was approved by the EU Parliament on 14th of April 2016 and was enforced on 25th May 2018. As the GDPR is a Regulation in European Union law, it has direct effect and does not need to be transposed into member states domestic legislation. The GDPR replaced the Data Protection Directive 95/46/EC and had initially been proposed to establish a uniform set of rules that would provide enhanced protection for citizens, foster innovation in the European Single Market and ensure that the European Union’s data protection law is fit for purpose in the digital age.452

The GDPR states that it objectives are to (1) lay down rules relating to the protection of natural persons with regard to the processing of personal data; (2) protect fundamental rights and freedoms of natural persons and in particular to the protection of their personal data; and (3) ensure free movement of personal data within the union with respect to those rights.453 The material scope of the GDPR is that it ‘applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system.’454 Data processed for the purposes of law enforcement falls outside the scope of the GDPR and under the Directive 2016/680, known as the Law Enforcement Directive. The Law Enforcement Directive must be transposed into domestic legislation. The territorial scope of the GDPR is explicitly global as Article 3 maintains that its requirements do not just apply to data controllers or

454 GDPR Article 2(1)
companies that are headquartered in the EU, but rather to any companies EU residents’ personal data.\textsuperscript{455} Overall, the GDPR ensures uniformity of data protection laws in the EU and covers a much wider scope than any other data-related legislation globally.

Given that the purpose of this project is to analyze the impact of automated processing systems on due process rights, and that the systems utilized by judges in their decision-making often focus heavily on profiling offenders to determine recidivism, this legislation report will focus on how the GDPR addresses profiling, and how effectively the Articles within the GDPR can safeguard individuals from decisions made based on automated systems. The specific articles of the GDPR that will be analyzed for the purposes of this legislative report include the highly topical Article 22 and its supporting Articles 13, 14, 15. Article 22 will be analyzed as it provides a legal protection from being subjected to a decision solely based on automated processing. Furthermore, Articles 13 through 15 will be analyzed as they appear to limit and determine what information data subjects receive in regards to the logic involved with automated decision making processes. First, however, some background analysis and the genesis of the GDPR and the advisory bodies will be explained.

\textbf{Background}

There are a number of key differences between the old Directive 95/46/EC (DPD) and the GDPR. One of the key differences is the new definition of personal data. Previously, the DPD defined personal data as ‘any information relating to an identifiable natural person; directly or indirectly by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.’\textsuperscript{456} In contrast, the GDPR defines personal data as ‘any information relating to an identifiable person… directly or indirectly, in particular reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more

\textsuperscript{455} GDPR Article 3
\textsuperscript{456} Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data Article 2(a) Repealed by GDPR (Hereinafter DPD)
factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.\textsuperscript{457} This is important as profiling an individual or their preferences using identifiers such as browsing history or biometric data will not longer be acceptable under the GDPR, unless the individual has consented, or is subject to a specific exception.

Another key difference is the uniform penalties enforced by the GDPR. Previously, under the DPD, EU Member states were free to adopt different notification laws when data breaches occurred.\textsuperscript{458} The GDPR imposes a requirement wherein data controllers must notify their supervisory authority and individuals affected by a personal data breach within 72 hours of learning about the breach.\textsuperscript{459} Although not overly relevant to the issue of profiling, it is a step forward for the protection of privacy rights. Violations of the GDPR come with increased penalties under Article 83, and Article 30 requires that records be kept of all processing activities, which must be made available to authorities upon request to enhance transparency.\textsuperscript{460}

The GDPR also established a new advisory body known as the European Data Protection Board (EDPB) under Article 68. The EDPB is composed of the representatives of national protection authorities, and the European Data Protection Supervisor (EDPS). The duties of the EDPB are to ensure the consistent application of the GDPR, fulfill advisory duties laid out under Article 70 and issue guidelines.\textsuperscript{461} The EDPB replaced the Article 29 Working Party (A29WP). The composition and purpose of the A29WP were set out under Article 29 of the DPD. The A29WP did, however, produce guidelines on automated decision-making and Profiling for the purposes of GDPR,\textsuperscript{462} which have been important in the Regulation’s interpretation and in academic literature.\textsuperscript{463}

\begin{footnotesize}
\textsuperscript{457} GDPR Article 4(1)
\textsuperscript{458} DPD Article 22
\textsuperscript{459} GDPR Article 33
\textsuperscript{460} GDPR Article 83 and Article 30
\textsuperscript{461} GDPR Article 70
\textsuperscript{462} Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679,’ 2017
\textsuperscript{463} See for example: Michael Veale & Lilian Edwards, ‘Clarity, surprises and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling,’ 2018
\end{footnotesize}
Profiling, Automated Decision Making & the Right to an Explanation

‘Machine learning (ML) or artificial intelligence (AI) systems are among the possible modes of decision-making, uniquely in danger of defying human understanding, and without human intervention or understanding would seem to flout the European ideas of autonomy and personhood.’ The provisions in the GDPR exist to provide some meaningful information to data subjects about how their data is used and many hoped in anticipation that the GDPR would provide residents with the ‘right to explanation’ and a form of protection from ‘profiling.’ The right to an explanation is viewed as a promising mechanism by both government and industry in the pursuit for accountability and transparency in algorithms, artificial intelligence, robotics, and automated systems.

The GDPR specifically addresses profiling and automated individual decision-making. ‘Advances in technology and the capabilities of big data analytics, artificial intelligence and machine learning have made it significantly easier to create profiles and make automated decisions with the potential to significantly impact an individual’s rights and freedoms.’ These processes are often very opaque and profiling can perpetuate existing stereotypes and social segregation leading to inaccurate predictions and unjustified discrimination in some cases. The GDPR defines profiling in Article 4(4) as:

---

464 Andrew D. Selbst & Julia Powles, ‘Meaningful information and the right to explanation,’ 2017 *International Data Privacy Law* 7(4) 223-242 p. 223
465 See for example: Sandra Wachter, Brent Mittelstadt & Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,’ 2018 *International Data Privacy Law* 7(2) 76-99; Selbst & Powels (n 13)
467 Article 29 Data Protection Working Party (n 462) p. 5
468 Article 29 Working Party (n 462) p. 5
'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, locations or movements;'

‘Profiling can be described as the process of gathering information and personal data, combining this individualized data with other (eg, personal, factual, statistical) data and analysing it through algorithms with the aim of predicting a person’s future conditions, decisions or behaviour.' In contrast, automated decision-making has a different scope that partially overlaps with profiling in some instances. Solely automated decision-making is the ability to make a decision by technological means without human intervention and automated decisions can be made with or without profiling. This is where Article 22 of the GDPR becomes relevant. Article 22 derived from article 15.1 of the DPD, but adopts a much broader approach, which is not limited to profiling, but it also includes other systems such as decision support systems. Article 22(1) states that:

‘the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.'

Article 22

Essentially Article 22 would put a general prohibition on decisions made about individuals on a solely automated basis. For example, Article 22 of the GDPR would

---

469 GDPR Article 4(4)
471 Article 29 Working Party (n 462)
472 Antoni Roig, ‘Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR),’ 2017 European Journal of Law and Technology
473 GDPR Article 22
hypothetically ban a banking institution from applying a solely online decision-making program to award loans to individuals. There are however, many exceptions to the right provided under Article 22. The right is excluded if the decision is necessary for a contract, or based on consent, or specifically authorized by a Member State for legitimate purposes. Further, if the decision is based on “special categories” of personal data (defined in Art 9 of the GDPR including race, religion and health), then automated decision-making is only allowed on the basis of explicit consent or substantial public interest.474

One of the primary safeguards that Article 22 provides would be that it requires that a human being is involved in the decision-making process under normal circumstances, outside of a few exceptions, which are listed in the GDPR. According to the Article 29 Data Protection Working Party, one cannot fabricate human involvement under Article 22 of the GDPR and the controller must ensure that oversight of the decision is meaningful, and that it should be carried out by someone who has the authority and competence to make the decision and that individual should consider all of the available input and output data.475 This appears to be an attempt to curb any potential ‘rubber stamping’ of an algorithmically prepared decision.476 It has been argued that ‘even when a human being formally makes a decision, for instance the decision to remove certain content from a social media platform, the human being may often be led to ‘rubber stamp’ an algorithmically prepared decision, not having time, context or the skills to make an adequate decision in the individual case.’477 Thus, in practice, the boundaries between fully automated processing and semi-automated processing are blurred.478

474 Michael Veale & Lilian Edwards, ‘Clarity, surprises and further questions in the Article 29 Working Party draft guidance on automated decision-making and profiling,’ 2018 Computer Law & Security Review 34 (2) 298-404
475 Article 29 Working Party (n 462)
476 Committee of Experts on Internet Intermediaries (MSI-NET), ‘Algorithms and Human Rights: study on the human rights dimensions of automated data processing techniques and possible regulatory impacts,’ 2017 Council of Europe p. 8
477 Ibid.
478 Ibid.
Article 22 can be criticized for having a limited scope in regards to the blurred line between semi-automated and solely automated processing. It only applies when the processing has been ‘solely’ by automated means, and various systems that have a significant impact on people’s lives are usually not considered ‘fully automated.’ For example, Article 22 would be excluded from many well-known algorithmic decisions on criminal justice risk assessments as the tools are often used as ‘decision support’ rather than the sole decision factor. Moreover, it is difficult to think of automated systems where significant decisions are made ‘solely’ by algorithms, as there usually is a ‘human in the loop’ at some point in the decision-making process. The attempt of the A29WP to include ‘meaningful’ human contribution in their guidelines demonstrates the right intention to curb rubber-stamping, however, they failed to define the term ‘meaningful.’ If any human involvement at all were allowed, through literal interpretation, to exclude a system from the ambit of Article 22, then its reach would be incredibly small. There is also a lack of enforcement mechanism within the GDPR as a whole to prevent a system from inadvertently becoming wholly automated through rubber-stamping.

Another issue with the definition under Article 22 is that it requires there to have been a ‘decision’, which produces ‘legal effects, concerning him or her, or significantly affects him or her.’ The GDPR does not define what a ‘decision’ actually is beyond the brief statement that it ‘may include a measure.’ Creators of these automated processing systems could argue that the output of a processing system is merely something to be used to assist a human in making a decision, and that on their own, these systems are incapable of processing the information and relevant uncertainties into an actionable decision.

Further, the definition under Article 22 also leaves uncertainties about what is considered to ‘significantly affect’ someone. To provide a contextual example, in 2013, a researcher investigating the delivery of targeted advertisements by Google AdSense using a sample of names associated with race and found that names associated with a minority

---

479 Veale & Edwards (n 474) p. 400
480 Lilian Edwards & Michael Veale, ‘Slave to the algorithm? Why a ‘Right to Explanation’ is probably not the remedy you are looking for,’ 2017 Duke Law & Technology Review
racial origin generated a much higher percentage of advertisements associated with arrest when compared to ads delivered to ‘Caucasian’ first names. This is a rare example of a ‘solely’ automated decision, which would pass the first hurdle of the Article 22 definition, but would the output of this automated processing system produce ‘significant’ or ‘legal’ affects against an individual? It is arguable that the mere presentation of this ad would not generate significant consequences, as there is no requirement to become a consumer of the services offered by this ad, and not even a requirement to click on the ad.

More recent examples include what the media has called the ‘Cambridge Analytica Scandal.’ This scandal involved Cambridge Analytical allegedly using data from Facebook to psychologically profile users to target them with material to help with both the Trump 2016 presidential campaign and the Brexit Campaign. While it has been demonstrated that there was certainly a breach in regards to lack of consent in the processing of individual’s data, which would fall under other provisions of the GDPR, the ‘significant affect’ requirement under Article 22 may remain unfulfilled. Facebook users are not required to act on the targeted material presented to them, and thus, may not have voted in either campaign, or even voted in opposition to what they were presented with. The wording of Article 22 has the potential to severely limit its scope, resulting in a major lack of protection for individuals subjected to automated and particularly semi-automated decision making.

---

481 Latanya Sweeney, ‘Discrimination in Online Ad Delivery,’ 2013 ACM Digital Library 11(3)
The Right to an Explanation

Prior to the implementation of the GDPR, many had hoped that it would make provisions to provide individuals with a ‘right to explanation’ in regard to both the automated data processing itself and how the processing reaches its ‘decision’ or output. Prior to the GDPR implementation, the Oxford Internet Institute forwarded that ‘the rapid spread of automated decision-making into sensitive areas of life, such as health insurance, credit scoring or recruiting, demands that we do better in allowing people to understand how their lives are being shaped by algorithms.’\footnote{Oxford Internet Institute, ‘submission to the Committee as part of the inquiry on algorithms in decision making,’ 2017 House of Commons’ Science and Technology Committee accessed March 15, 2019 from: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/69003.pdf} IBM thought it was important that explanations uncover how algorithms ‘interpreted their input’ as well as ‘why they recommend a particular output,’\footnote{International Business Machines (IBM), ‘written evidence submitted by IBM’ 2017 House of Commons Science and Technology Committee accessed March 15 2019 from: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/71691.pdf} and the University College London wanted a ‘meaningful right to explanation strengthened to include ‘semi-automated’ as well as the automated decisions that are covered by the GDPR.’\footnote{University College London, ‘Written evidence submitted by UCL (ALG0050),’ 2017 House of Commons Science and Technology Committee accessed March 15 2019 from: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/69093.pdf}

If the aforementioned reasons weren’t enough, another reason for the requirement of a ‘right to explanation’ within the GDPR would be that providing individuals with meaningful information would help them to contest decisions when an adverse or otherwise undesired decision is received. The right to contest decisions is provided as a safeguard by Article 22(3) against automated decision-making, however, whether a right to explanation exists within the GDPR is highly contested.
Recital 71 of the GDPR states that ‘the data subject should have the right … to obtain an explanation of the decision reached.’ Recitals, however, are not legally binding and the right to explanation is not mentioned in the binding texts of the article, including Article 22, as it was removed during the legislative process. According to some academics, the purposeful omission of this text from Article 22 may not be an oversight but suggests that legislators did not intend to implement a right to explanation of specific decisions within the GDPR. The Article 29 Data Protection Working Party has attempted to preserve the much desired ‘right to explanation’ by noting that ‘it is good practice to ensure that information about profiling is not only easily accessible for a data subject but that it is brought to their attention, whether the processing falls within the narrow scope of Article 22(1) or not.’ Despite the Article 29 Working Party’s attempts, the UK’s House of Commons concluded that the GDPR provides no ‘wider right to explanation’ for UK citizens than the previous legislation, and the Digital Economy minister of France, Mounir Mahjoubi recently stated that its government should not use any algorithm whose decisions cannot be explained. It been argued, however, that a right to explanation may exist through Articles 13 through 15.

**Articles 13-15**

Together, Articles 13 through 15 have been called the ‘Magna Carta’ of data subject’s rights to obtain information about the data held about them and to scrutinize the legitimacy of processing that data. Articles 13-14 create notification duties for data controllers, while Article 15 establishes a corresponding right of access for data subjects. The guidelines state that Article 15(1)(h) is seen to provide identical information as

---

486 Recital 71, *EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679*

487 Wachter, Mittelstadt, Floridi (n 466) p. 6

488 Ibid.

489 Ibid.


491 Wachter, Mittelstadt, Floridi (n 466) citing Florian Schmidt-Wudy, ‘DS-GVO Art. 15 Auskunftsrecht der betroffenen Person’ in Heinrich A. Wolff and Stefan Brink (eds), Datenschutz-Grundverordnung (18th edn, beck-online 2016) Rn 2
Article 12(2)(f) and 14(2)(g) in the DPD which requires the data controller to ‘provide the data subject with information about the envisaged consequences of processing, rather than an explanation of a particular decision.’ Although incredibly similar, the right of access must be invoked by the data subject, whereas the notification duties are placed upon the data controllers. Specifically, Articles 13-14 provide the right to be informed about: (1) the existence of automated decision-making under Article 22(1) and (4); meaningful information about the logic involved; and (3) the significance and envisaged consequences of such processing. Articles 13-15 also describe the type information needs to be provided if data is collected, either immediately when collected from the data subject (Art 13), the latest after a month when collected from a third party (Art 14), or at any time if requested from the data subject (Art 15).

Article 12(7) clarifies that the aim of Articles 13 and 14 is to provide ‘in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing.’ The two requirements stemming from these Articles are 1) that the notification occurs prior to processing and 2) that the information provided must be meaningful to its audience. Currently, broadly applicable information, rather than a personalized disclosure appears to fulfill the requirement of providing ‘meaningful information’ or a ‘meaningful overview.’ ‘Legal scholars have suggested that notification duties can be satisfied via updates to existing privacy statements or notices.’ This position is further supported by the Working Party’s guidelines on transparency, which state that the notification duties under Articles 13-14 can be satisfied via standardized

---

493 Ibid.
494 Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679,’ 2017 p. 31
495 Wachter, Mittlestadt, & Russel, ‘Counterfactual explanations’ (n 492)
496 Article 12(7)
497 Wachter, Mittlestadt, & Russel, ‘Counterfactual explanations’ (n 492) p. 19
privacy notes, visualization tools, and icons.\textsuperscript{498} While these notifications may be helpful in the context of something such as social media use, updates to existing privacy statements may not work in all circumstances. For example, data subjects often do not have access to the algorithms used to process whether they will be approved for a loan or not. In the context of risk assessment instruments in criminal justice, a pop up notification would hardly suffice. Perhaps verbal explanations or formal written notifications with the assistance of an expert should be used in these circumstances.

This type of information provided under Articles 13-15 may assist individuals in contesting automated decisions made against them, but the existence of a ‘right to explanation’ following the output of an automated decision appears to be quite unconvincing. Each disclosure under Articles 13 and 14 must occur prior to data processing; otherwise the right not to be subject to automated decision-making could never be realized. However, according to the Article 29 Working Party, the process of decision-making and the algorithm itself does not need to be fully disclosed, but rather a description of the logic of the algorithm, which may include a list of data sources or variables.\textsuperscript{499} Moreover, the Article 29 Working Party notes that disclosure need not be complex, but rather explained in a way that makes sense to the data subject.\textsuperscript{500} Whether this actually happens is up for debate. Arguably, Articles 13 and 14 would still not provide a ‘right to explanation’ of an algorithmic output as the data controller appears to be only required to report the ‘envisaged consequences,’ before the decision is made, and not necessarily an explanation after a decision has been made. Moreover, a description of the logic of an algorithm may not provide an individual with enough information to contest and automated decision.

\textsuperscript{499} Ibid.
\textsuperscript{500} Ibid. p.14
Overall Impact of the GDPR

The GDPR provides an enhanced definition of ‘personal data’ and legislates harsher penalties for those who breach the Articles under the Regulation. The Regulation’s wide territorial and material scope provide for uniformity amongst Member States, and its implementation marks a step in the right direction for data protection. In regards to providing protection from profiling, Article 22 does provide data subjects with the right not to be subjected to automated processing that produces significant or legal affects, but only on the grounds that this processing is ‘solely’ automated. The language within Article 22 is vague and will exclude a large number of individuals from accessing the right due to ‘rubber stamping’ or brief interaction by nominal humans.

There was hope that the GDPR would provide a greater understanding of how automated processing works and provide a right to explanation. The GDPR certainly acknowledges this right through Recital 71; it lacks clear-cut provisions that provide for any formal explanation to an individual. While the notification duties are a step in the right direction, going as far to say the GDPR provides for an explicit right to explanation may be dishonest.
Bibliography

Primary Sources:

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data


Secondary Sources:

Andrew D. Selbst & Julia Powels, ‘Meaningful information and the right to explanation,’ 2017 International Data Privacy Law 7(4) 233-242


Edwards, L. and Veale, M., ‘Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy you are Looking For,’ 2017 *Duke Law and Technology Review* 16(1) 1-65


Sweeney, L., ‘Discrimination in Online Ad Delivery,’ 2013 *ACM Digital Library* 11(3) 1-19

University College London, ‘Written evidence submitted by UCL (ALG0050),’ 2017 *House of Commons Science and Technology Committee* accessed March 15 2019


Wong, J.C., ‘The Cambridge Analytica Scandal Changed the World – but it didn’t change Facebook,’ 2019 *The Guardian*

Wong, J.C., ‘The Cambridge Analytica Scandal Changed the World – but it didn’t change Facebook,’ 2019 *The Guardian*
Appendix 3: Oswald et al. ALGO-CARE

The following images are of the framework known as ALGO-CARE developed by Oswald, M., Grace, J., Urwin, S. and Barnes, G.C., ‘Algorithmic risk assessment policing models: lessons from the Durham HART model and ‘Experimental proportionality,’ 2018 Information & Communications Technology Law 27(2), 223-250

Disclaimer: These images are attached for references purposes only and I do not claim them to be my own.
<table>
<thead>
<tr>
<th>A</th>
<th>Advisory</th>
<th>Is the assessment made by the algorithm used in an advisory capacity? Does a human officer retain decision-making discretion? What other decision-making by human officers will add objectivity to the decisions (partly) based on the algorithm?</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>Lawful</td>
<td>On a case-by-case basis, what is the policing purpose justifying the use of algorithm, both its means and ends? Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes? In what way will the tool improve the current system and is this demonstrable? Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim? Is the operation of the tool compliant with national guidance?</td>
</tr>
<tr>
<td>G</td>
<td>Granularity</td>
<td>Does the algorithm make suggestions at a sufficient level of detail/granularity, given the purpose of the algorithm and the nature of the data processed? Is data categorised to avoid ‘broad-brush’ grouping and results, and therefore issues potential bias? Do the benefits outweigh any technological or data quality uncertainties or gaps? Is the provenance and quality of the data sufficiently sound? Consider how often the data should be refreshed. If the tool takes a precautionary approach towards false negatives, consider the justifications for this.</td>
</tr>
<tr>
<td>O</td>
<td>Ownership</td>
<td>Who owns the algorithm and the data analysed? Does the force need rights to access, use and amend the source code and data analysed? How will the tool be maintained and updated? Are there any contractual or other restrictions which might limit accountability or evaluation? How is the operation of the algorithm kept secure?</td>
</tr>
<tr>
<td>C</td>
<td>Challengeable</td>
<td>What are the post-implementation oversight and audit mechanisms e.g. to identify any bias? Where an algorithmic tool informs criminal justice disposals, how are individuals notified of its use (as appropriate in the context of the tool’s operation and purpose)?</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A</td>
<td>Accuracy</td>
<td>Does the specification match the policing aim and decision policy? Can the stated accuracy of the algorithm be validated reasonably periodically? Can the percentage of false positives/negatives be justified? How was this method chosen as opposed to other available methods? What are the consequences of inaccurate forecasts? Does this represent an acceptable risk (in terms of both likelihood and impact)? Is the algorithmic tool deployed by those with appropriate expertise?</td>
</tr>
<tr>
<td>R</td>
<td>Responsible</td>
<td>Would the operation of the algorithm be considered fair? Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review alongside other IT developments in policing? Would it be considered to be for the public interest and ethical?</td>
</tr>
<tr>
<td>E</td>
<td>Explainable</td>
<td>Is appropriate information available about the decision-making rule(s) and the impact that each factor has on the final score or outcome (in a similar way to a gravity matrix)? Is the force able to access and deploy a data science expert to explain and justify the algorithmic tool (in a similar way to an expert forensic pathologist)?</td>
</tr>
</tbody>
</table>

*Or as Braunies and Goodman put it, what is the ‘predictive goal’? n52 (51).*
The Algorithms in Policing – Take ALGO-CARE™ framework is intended to provide guidance for the
use of risk-assessment, predictive, forecasting, classification, decision-making and assistive policing
tools which incorporate algorithmic machine learning methods and which may impact individuals
on a micro or macro level

| A | Advisory | Care should be taken to ensure that an algorithm is not inappropriately
|   |          | fettering an officer’s discretion, as natural justice and procedural fairness
|   |          | claims may well arise. Consider if supposedly advisory algorithmic
|   |          | assessments are in practice having undue influence. If it is proposed that an
|   |          | algorithmic decision be automated and determinative, is this justified by the
|   |          | factors below? Data protection rights in regard to automated decisions may
|   |          | then apply.

| L | Lawful   | The algorithm’s proposed functions, application, individual effect and use of
datasets (police-held data and third party data) should be considered against
necessity, proportionality and data minimisation principles, in order to
inform a ‘go/no-go’ decision. In relation to tools that may inform criminal
justice disposals, regard should be given to the duty to give reasons.

| G | Granularity | Consideration should be given to common problems in data analysis, such as
those relating to the meaning of data, compatibility of data from disparate
sources, missing data and inferencing. Do forces know how much averaging
or blurring has already been applied to inputs (e.g. postcode area averages)?

| O | Ownership  | Consider intellectual property ownership, maintenance of the tool and
whether open source algorithms should be the default.¹ When drafting
procurement contracts with third party software suppliers (commercial or
academic), require disclosure of the algorithmic workings in a way that
would facilitate investigation by a third party in an adversarial context if
necessary. Ensure the force has appropriate rights to use, amend and
disclose the tool and any third party data. Require the supplier to provide an
‘expert’ witness/evidence of the tool’s operation if required by the force.²
<table>
<thead>
<tr>
<th>Column</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Challengeable</td>
<td>The results of the analysis should be applied in the context of appropriate professional codes and regulations. Consider whether the application of the algorithm requires information to be given to the individual and/or legal advisor. Regular validation and recalibration of the system should be based on publicly observable (unless non-disclosable for policing/national security reasons) scoring rules.</td>
</tr>
<tr>
<td>A</td>
<td>Accuracy</td>
<td>How are results checked for accuracy, and how is historic accuracy fed back into the algorithm for the future? Can forces understand how inaccurate or out-of-date input data affects the result?</td>
</tr>
<tr>
<td>R</td>
<td>Responsible</td>
<td>It is recommended that ethical considerations, such as consideration of the public good and moral principles (so spanning wider concerns than legal compliance) are factored into the deployment decision-making process. Administrative arrangements such as an ethical review committee incorporating independent members could be established for such a purpose (such as Cleveland &amp; Durham Joint External Ethics Committee or the National Statistician’s Data Ethics Advisory Committee).³</td>
</tr>
<tr>
<td>E</td>
<td>Explainable</td>
<td>The latest methods of interpretable and accountable machine learning systems should be considered and incorporated into the specification as appropriate.⁶ This is particularly important if considering deployment of ‘black box’ algorithms, where inputs and outputs are viewable but internal workings are opaque (the rule emerges from the data analysis undertaken). Has the relevant Policing &amp; Crime Commissioner been briefed appropriately?</td>
</tr>
</tbody>
</table>