To what extent is the use of recidivism prediction and sentencing software in the criminal justice system justified from an ethical perspective?

Term Paper

by

Name	Faizan Zafar
Study Program	Master of Data Science
Semester	Summer Semester 2019
Email	faizan.zafar@rwth-aachen.de

In Fulfilment

Of the Requirements for the course

Title	Ethics, Technology and Data
Number	71.00002
Professor	Nagel, Saskia Kathi, Univ.-Prof. Dr.
Email	saskia.nagel@humtec.rwth-aachen.de

RWTHAACHEN
UNIVERSITY

(2019)

Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Author	Faizan Zafar
Signature	

Table of Contents

Glossary of Terms

AI	Artificial Intelligence
RPS	Recidivism Prediction Software
SSS	Sentencing Software System
CJS	Criminal Justice System
PSI	Presentence Investigation

ABSTRACT

Artificial Intelligence encompasses a broad spectrum of capabilities of a computing system to interpret and react to its environment autonomously (with minimal to no human intervention) and perform meaningful tasks that would normally rely on human insight. Expanded automation with intelligent machines offers opportunities for governments and industries to progressively complete complex tasks with a degree of enhanced productivity and finesse far beyond that of humans, producing considerable social and economic benefits simultaneously.

One of the most recent and revolutionary developments of the criminal justice system has been the introduction of advanced software programs designed to assist judges and prosecutors, in assessing the risk of criminal offenders and additionally aiding the court staff in sentencing these offenders appropriately. These programs employ complex algorithms, that make use of several social, economic, geographic and demographic factors to reach a supposed prediction measure of an individual's criminal risk, which later forms the basis or constitutes as a single or even fundamental factor in drafting a sentencing decision.

These intelligent programs have the potential to be a permanent part of the criminal justice system, providing investigative assistance and accurate sentencing, and allowing criminal justice professionals to better maintain public safety. However, there arises the need for crucial ethical decisions as this technology is integrated into the current system, primarily due to the vast diversity of criminal cases and critical sensitivity of legal matters.

This paper targets the ethical implications and arguments revolving around the use of recidivism prediction tools and sentencing software (commonly existing as a singular entity), in the criminal justice system. I present my normative stance, in correlation with the literature, of why I consider it immoral for the legal system to be influenced by these highly complex algorithms, whose reasoning process is out of the bounds of human rationale. I argue about the horrific prospect of algorithmic biases in recidivism prediction and sentencing in criminal proceedings. I debate on the issue that these sophisticated algorithms give birth to moral biases. I go on to present my arguments on a specific application area pertaining to the proceedings of homicide cases. Additionally, I dive into the legal and public questions raised by the use of risk assessment software in sentencing decisions.

This paper touches on various aspects in the vast domains of machine bias, fairness in AI and black-box algorithms.

1. INTRODUCTION

1.1   Background

Before the integration of technology in the CJS, the tasks associated with dealing with criminals mainly rested with police officers, lawyers, jury, judges and other criminal justice professionals, who gained expertise in their respective areas through rigorous practice over the course of several years. However, limitations of the human workforce and increasing complexity of legal matters allowed space for the latest technological trends to sweep in. The advancement of AI in the judicial and legal systems came to the rescue; the prospect of a machine, able to perform numerous computations in a fraction of a second, and conform with legislations and sentencing policies, definitely proved to be a gamechanger. These machines have the capabilities to detect anomalous patterns and irregular behaviors, conserve judicial infrastructure and decipher criminal networks among several other useful applications.

Predictive analysis and decision-making in the legal system are often a set of complex tasks that involve collecting huge amounts of relevant information (usually criminal records), analyzing formal laws and precedents, as well as assessing feasible options. The output of this analysis is used to formulate potential outcomes in terms of risk assessment or suggesting sentences. Similarly, AI is being used to predict potential culprits of violent crimes such as homicide, based on personal features, demographic traits and geographic associations. Simply put, these advanced algorithms influence a human decision maker in both; giving out normative decisions and guiding them towards a certain supposedly legitimate direction.

Risk assessment tools and sentencing software (commonly existing as a singular entity), most of which incorporate machine learning algorithms, are being used in a variety of contexts within the legal system including but not limited to; prison rehabilitation programs, pre-trial risk assessment, and sentencing. They are able to generate risk models based on vast quantities of datasets. They can also formulate risk scores and sentencing decisions, based on data of prior offenses.

A significant issue surrounding algorithmic decision-making is that the software itself is essentially a *black-box* of extraordinarily complex decision models involving hundreds of features and millions of data points. Complexity in these algorithms creates opportunities for bias to inadvertently affect their targets; in this case, the alleged offenders (or defendants). While the court staff and defendant can control what data goes into these systems, and programmers can instruct them how to weigh in different variables, it is extremely challenging (even nearly impossible) to program these systems to justify their decisions. Consequently, I regard these algorithms as impenetrable black boxes that defy scrutiny.

Drawing a sentencing decision mainly involves two aspects; deciding how to punish an alleged culprit and, if a judge chooses imprisonment, the length of the sentence. Determining the severity and length of the punishment often draws upon a number of different theories of punishment, including individual retribution, rehabilitation, deterrence, and incapacitation [1]. Judges often base their decisions on multiple theories, despite their varied goals [1].

An important question surrounding sentencing law is how legislators should strike an appropriate balance between consistency and individualization in punishments. I think it is essential to prioritize individualization as it enhances segregated punishment, which increases legislators' moral objective to pursue optimal levels of fairness and remediate biases. I understand that AI-based solutions terribly fail to highlight the individualization aspect but help to maximize consistency of the sentencing process.

Intentional homicide caused the deaths of almost half a million people (437,000) across the world in 2012 [7]. Beyond that, this form of violent crime has a broad impact across all societies. Globally, police forces are able to identify and arrest one or several suspects for a particular homicide in a little over 60% of cases [7], allowing the case to be turned over to the prosecution service. An important indication of the criminal justice response to homicide is the conviction rate, which leads to 43% perpetrators being convicted [7].

With legal proceedings of such intense nature, especially in homicide cases, where there are numerous variables at play, even experienced judges have a hard time taking decisive action. A commonly adopted approach is data collection and analysis of the alleged murderer's actions at each stage of the prosecution, so the court staff can make an informed decision. However, with corrupted datasets possessing societal biases and murder cases involving a higher degree of legal and ethical complexity, it is unlikely that AI can deliver what it promises; fully automated decision-making that is fair to the offender. Hence, there arises the need for the implementation of a logical punishment scheme for the culprit, delivered in the most efficient way possible. This action must be aimed at consequently bringing justice to the victim.

1.2   Thesis Statement

In this paper, I argue that it is morally wrong for us, as rational individuals with the power of reason and moral observation, to exercise the decision of SSS and consequently let its underlying black-box algorithms decide the fate of a human being. This decision might be irreversible in nature, especially in the case of homicide criminals. I form my thesis in the frameworks of risk assessment and sentencing. In court affairs, AI is capable of predicting criminal recidivism and this prediction measure is actively used in the formation of a sentencing decision, so both these aspects go hand in hand. I deem it as quite inhuman and even brutal to use a computing system for determining the ability of a convicted individual to indulge again in illegal or immoral activities.

In sentencing, judges use risk assessment software with the goal of making sentences more uniform and predictable, while remaining sensitive to public safety concerns. Another reason

for their employment is to identify low-risk felons, so as to be able to impose non-carceral punishments [3]. A variety of states in the United States of America employ a risk assessment tool called COMPAS [6], which provides judges with a risk assessment score for defendants, based on the defendants' answers to a comprehensive series of questions. Moreover, COMPAS deems the defendant's profile as high, medium or low threat and assigns him the respective category. Majority of the RPS and SSS function in this way. In this paper, I refer to this tool and criticize its usage, mainly because of its widespread and unprecedented use. COMPAS has also been publicly covered by ProPublica.

1.3   Motivation

As the influence and scope of RPS increase, journalists and legislators have raised concerns that the statistical models from which they are derived might unintentionally contain human biases. With this paper, I criticize the affects that the application of this software has on humans. I am of the view that recidivism predictions and sentencing decisions (or any substantial legal decisions) made by AI must possess (in practice) the following criteria to be considered *moral* (or at least to a close definition of it).

- **Logic:** decisions must conform to concrete logic within human rationality.

- **Individualization:** numerous factors specific to the circumstances of the offense and offender are extremely relevant to ensuring complete justice to the victim.

- **Coherence:** decisive factors must be given substantial weightage in a decision, with each part of the final decision connecting in a natural flow.

Predictive algorithms function perfectly in summarizing all relevant information of a criminal case in a more efficient way than the human brain. But are they able to rationalize their decision-making processes? According to literature, actuarial assessments do a better job than individualized judgment at predicting risk and therefore help exceptionally in reducing imprisonment and promoting public safety. While this might be occasionally true for generic cases or low priority felons, actuarial assessment does not favor people of color, specific ethnicities or races, individuals with track record of minor offenses, juveniles, serious offenders aiming to transform their life around and homicide criminals.

1.4   Essay Organization

This paper is organized in 6 chapters:

Chapter 1 consists of the background, thesis statement and motivation for the topic under discussion.

Chapter 2 consists of the logically compelling arguments supporting the main thesis.

Chapter 3 consists of the counterarguments to the supporting arguments of the thesis.

Chapter 4 consists of possible responses to the primary objections to the main arguments.

Chapter 5 includes a brief summary of the main arguments and a satisfactory conclusion for the paper.

Chapter 6 consists of the complete bibliography.

2. ARGUMENTS

Introduction

In this section, I present the arguments in light of the existing knowledge related to the subject of inquiry. I also highlight the relationship of the research with examples for purposes of good representation and critical review of the existing literature. I have cited various academic articles and presented my findings in correlation with them.

2.1   Arguments

Because of biased policing patterns, racism in sentencing and influenced criminal assessments, (many of which are reinforced by RPS) the minorities, poor and people of colour face an unbalanced measure of legal actions with respect to previously witnessed violations via the software. I think the source of this problem is overlooked, more often than not i.e. the nature of the prediction itself. The algorithms within an RPS act blatantly like a mirror as they function on the principle of learning from the past because the future is bound to repeat it, while neglecting outliers, anomalous cases or highly individual features. This basically means that a collection of traits that are correlated with crime execution in the past will definitely correlate with crime execution in the future, which is not necessarily true in practical situations. A similar interpretation is that RPS extracts patterns from the past dataset and then analyses projections about the future. Mapping this issue to a homicide case, with fixed sentences for most murder types, the biased algorithm can either sentence or release someone erroneously, leaving the court staff none the wiser. Additionally, considering the case of the State of Wisconsin, USA vs. Eric Loomis [5], it is clear that COMPAS violated Loomis' right to an individualized sentence because it relied on information about the characteristics of a larger group to make an inference about his personal likelihood to commit future crimes [1].

The court staff is able to distance themselves from any morally suspicious decisive action of RPS by attributing it solely to the algorithm. They defy to weigh in the emotions of a defendant (and his supporters) when making a recidivism prediction (or even sentencing decision) and eventually, it just becomes a game of number crunching and being statistically accurate. The complexity and automated nature of RPS are immediately blamed so as to explain why it arrives at a morally inept conclusion. This wrongfully allows the court staff to imply that the result is unintended and something which they are not responsible for. This prospect is quite harmful for alleged murderers as the software can decide between parole sentence, life imprisonment or death penalty. From the general public's perspective, the judge merely acts as a liaison between the software and criminal offender. While this may truly be not the case yet, it is what AI enthusiasts are rooting for. Holding algorithms accountable is a different domain entirely but, I believe, the ability of a human decision-maker to introduce an element of humanity and emotional understanding in these tough circumstances will always be of utmost

importance.

It is quite vague as to how a judge should employ a risk assessment score to inform a final sentencing decision as compared with a decision in the pre-trial phase of a criminal proceeding. The objective in sentencing is to punish and possibly deter the convicted individual and promote his rehabilitation. It is regarded as the most difficult and crucial time for a judge in a criminal prosecution. Due to this burden, a judge might be attracted to using this score generated by RPS. Interestingly, in the pre-trial phase, a judge faces a relatively simple decision i.e. should the criminal stay in jail for the complete duration of the pre-trial phase or be allowed some time off with the family. Later, at the sentencing stage, the judge also has to determine the length of the punishment. Deciding upon this length may involve numerous factors including but not limited to; the offender's criminal history, nature and severity of the offense, offender's documented remorse, level of harm or injury incurred to the victim and offender's personal circumstances. For homicide offenses, additional factors such as the degree of murder, possibility of self-defence, liquidity for paying fines and associated damages are considered. These factors weigh in with different percentages into the final sentencing decision. Understandably, outsourcing this set of highly complex and human processes to a software seems irrational and impractical in the long run.

The algorithms within an RPS are not concerned with the fact that various factors that are proven to increase a risk score might also be interpreted as mitigating evidence. This is where contextual judgement within a criminal prosecution comes in and is critically important to understand before concocting assumptions, especially in the current era of evidence-based sentencing. Now, for example, a young defendant, hailing from a poor background and barely possessing educational credentials, is designated at a higher risk for recidivism according to RPS, but the very same characteristics can also diminish his culpability and rationalize a more lenient sentence. Similarly, examining the intention of murder, passion of crime, nature of manslaughter and contribution to assisted suicide or vehicular homicide, using AI, might tip the scales against, let's say, a minor with proven medical history of severe social anxiety disorder. Another issue is that these programs are accurate only in specific geographic areas in which they are deployed, however, they might be inaccurate for other areas. Defining features of a region such as crime rate, economic prosperity and neighborhood characteristics might out rule the personal circumstances of, let's say, a male offender of colour from the region. In these cases, deriving context from statistical evidence can prove to be brutal for the alleged offender. This also goes against the universally fundamental moral imperative to avoid pain at all costs.

From reviewing relevant literature and assessing the press release statements of technologically advanced courts (such as the Judicial System of the State of Wisconsin, USA), it becomes clear that decisions carried out by RPS are not understood by the judges who apply them. This is bound to happen with any black-box model. I believe these supposedly intelligent systems are analogous to all-knowing (even omniscient) experts that the CJS or general public are unable to argue with. This can be troublesome for felons who aim to make a plea deal or display remorse to the judge and jury for a reduced sentence and definitely dangerous for falsely accused offenders. Moreover, I think that general concerns of this dilemma are far worsened

due to the value people give to AI in the modern society. This leads to the prevalence of an unintended automation bias. Judges might be oriented towards an automated decision just because of its 'automated' nature and be highly likely to assume that quantitative methods are far superior to ordinary verbal reasoning. This prime form of automation bias can easily change a technological nudge into an authoritative decision, which is a frightening prospect for most offenders ranging from street thugs to convicted murderers. In this way, AI can work to anchor significant decisions towards an unknown certainty that is bound to be morally inappropriate.

The proprietary nature of RPS and SSS keeps their contents or functioning principles hidden from defendants and even courts. This legal protection renders it impossible to prove that software such as COMPAS actually produce results that are contrary to the evidence it receives, primarily because the algorithm is predictive. For instance, the algorithm can be said to conform with defined legislations when a defendant with a high-risk assessment score goes to jail and reoffends later in life. However, if the same defendant does not reoffend; the predictive outcome might have no correlation with the actual outcome. Possibly the severe punishments in prison worked for the better or any other factor changed, and the defendant was deterred from reoffending, thus achieving a positive outcome. Another issue is that this proprietary nature prohibits the defendant, his counsel and the general public from understanding the reasoning processes behind algorithmic decision-making. It heavily undermines the defendant's right to demand a re-trial or a reduced sentence. Since there is no proof of misconduct or malpractice while using RPS, as it is part of the State's legislation, the counsel cannot challenge its decision on solid grounds. As I see it, the CJS is taking a leap of faith with this software and 'playing with fire' since this proprietary nature does not provide a safe area of argument between the judge and the defendant. I think not having a proper legal framework with respect to the public and private features of the software adds to overall moral uncertainty.

An intimidating phenomenon is the excessive reliance on technology, to the extent that it affects the user and the target in an unintentional way. Eventually, the responsibility and independence of the judge are threatened and the reliable court staff may become dispensable. In the CJS, where each criminal proceeding encompasses its set of ethical complexities, AI can prove to be counterproductive. Another related issue is that sometimes there is a conflict of interest between different legal stakeholders and due to inherent biases, the SSS might be favorably aligned towards interests of a certain legal entity, racial ethnicity or demographic group, enabling its superiority over the human actors involved. Tech firms owning this software claim to develop AI systems by explicitly modelling the reasoning processes that judges use when sentencing offenders. However, they can also, indirectly, manipulate the naïve nature of the CJS and aim to profit some beneficiaries more than others. Upon review of different publicly available criminal justice guidelines, there seems to be no concrete legal literature in this particular area of AI-guided decision making and the existing informal rulings are quite ambiguous. Thus, making it reasonable as well as conceivable for crime ethicists and legislators to not be comfortable and reassuring while adopting AI in the near future.

3. COUNTERARGUMENTS

Introduction

In this section, I articulate counterarguments to the previous arguments to contradict my position regarding the topic under discussion. I highlight the opposing aspects so as to give a clear picture of the dilemmas surrounding the use of automated recidivism prediction and sentencing in the CJS.

3.1   Counterarguments

While subjective prediction only allows a bizarre reflection of analog data, AI delivers what it promises; statistically accurate prediction measures. The constituting dataset might contain biases which the algorithm builds upon, however, those are derived from human actions. There are no external influences or emotions involved, and hence the predictions have a higher degree of objectivity. Algorithmic recidivism prediction helps the CJS rely less on subjective intuition and make truly evidence-based decisions about who can safely be released or should be serving jail time, thus saving immense costs and reducing prison overcrowding. The development of computing systems that support court staff in determining appropriate sentences is generally motivated by a desire to ensure efficiency and consistency in decisions. These programs are able to calculate and organize statistical data derived from databases of prior decisions or retrieve similar cases using highly complex but accurate algorithms. For a court system which prioritizes consistency in decisions and prefers the use of AI for critical tasks, RPS can ensure efficiency and streamlining of human processes.

It is not in the inherent nature of an algorithm to be biased towards or against a certain person or racial ethnicity. A risk assessment score from COMPAS, for example, is based upon the input to its carefully crafted questionnaire and publicly available data about the defendant's criminal history. The judge or court staff do not have any influence over its inner functions, and eventually it boils down to the defendant's interaction with the RPS. Referring to the case of the State of Wisconsin, USA vs. Eric Loomis in 2016 [5], Mr. Eric Loomis pleaded guilty to crimes related to a drive-by shooting. The trial judge ordered a PSI, which in turn used COMPAS. He used the PSI and COMPAS reports in his decision to sentence Loomis in the maximal range for only the crimes to which he pled guilty. In this scenario, the court argued that he was fully responsible for the sentence, since most of the input information came from a questionnaire that he completed and from public records. Simply put, the algorithm interpreted this information and gave it numerical meaning with the purpose of making it useable and helpful, rather than harmful.

For the sake of argument, let us consider a hypothetical case where the risk assessment score is either the only factor or the determinative factor in a sentencing decision. For this scenario, it makes perfect sense to criticize the program's creators, condemn the governments who allow to deploy it, allocate resources to understand its functioning principles and advocate to open source its data and code. However, in practical situations, the risk assessment score is simply one piece of information among several others that the judge considers in a sentencing decision. So, the amazing benefits outweigh the supposed disadvantages. It is purely up to the judge and court staff, to decide upon the degree of dependability upon the system.

Looking beyond personal circumstances, I think the locality of a defendant contributes highly substantial weightage in influencing recidivism prediction or sentencing and reasonably so. For example, if region X is highlighted to have more homicide victims than region Y, it is safe to assume that an alleged murderer from region X might have actually been involved in the heinous crime. It is also perfectly reasonable to conclude that, let's say, a man (of color or not) hailing from region X, with some criminal background might have been involved in a relevant criminal activity, without giving much significance to individual features. In practical situations, locality and other external characteristics are evaluated by multiple legal experts and the final decision is then an average of individual decisions. Thus, the 'humanity' element is included in decisions even when RPS is used.

The black-box nature of algorithms within an RPS is what makes them accurate to a great extent. An AI system provides decision-support to a judge and court staff, while strengthening their competence in processing scattered knowledge. It reinforces their information management abilities with technology-based means. It is able to receive, save, utilize and display relevant knowledge regarding the justice decisions under scrutiny. Its capabilities are defined by in-built data pattern extraction mechanisms, the ways in which it can represent these complex patterns and give them meaning. Integrating AI into the inner workings of the legal system by enforcing regulatory measures through well-defined policies promotes automated decision making. With these amazing advancements and astounding benefits, it would be a huge misfortune for the CJS to miss out on RPS and SSS.

With the rise of machine learning powered applications, the AI Industry is striving towards highly innovative ventures. This industry works on the principles of any modern industry; creating demand, providing supply, maximizing profits, minimizing losses and allocating economic resources appropriately. It is definitely not among their prime or even secondary interests to focus upon the ethical consequences of the applications of this technology. Creators of COMPAS and similar corporations have a prominent interest in masking their products in secrecy, so as to remain competitive against other firms in the market. Moreover, these lucrative tech companies can afford to utilize the power of effective legal services to keep their algorithms away from the open source community.

AI research is applied to solve purely technical problems, rather than ethical ones. AI developers and data scientists are focused towards rigorously improving the out-of-sample accuracy of their machine learning models, no matter the costs involved. This idea is reinforced by companies who are aiming to outsmart their rivals. Modern research is being conducted to create more efficient implementations of AI for the legal system which aids in resolving difficult cases, with little to no regard for moral bias minimization.

4. RESPONSES TO COUNTERARGUMENTS

Introduction

In this section, I author legitimate responses to the previous counterarguments to provide insight into logical reasoning behind the views of the opposing stance.

4.1   Responses to Counterarguments

Referring to past datasets for similar cases seems to be an extraneous step because of the uniqueness of each offense and each offender. The primary criteria taken into account for deciding the length and severity of sentencing punishments should entirely be based on the motivation of providing justice to the victim. For this to be effective, only the alleged offender's culpability and the seriousness of the offense must be the primary determinants. External characteristics such as race, ethnicity, religion, age and gender should not be included among recidivism prediction and sentencing criteria as offenders have no control over them. Even with homicide cases, an array of psychological, financial, intimate and cultural motives of murder might out rule the said external characteristics. So, substantial efforts should be spent in deciphering between these and providing justice, rather than incorporating unfamiliar technology to do the same.

The defendant's access to RPS's inputs does allow him to have some opportunity to explain his position, however, it is insufficient to verify the RPS's accuracy. For instance, it is logically improper to draw a targeted conclusion about an alleged offender, derived primarily from characteristics pertaining to a specific (not necessarily related) group. Understandably, ensuring factual correctness of inputs to the RPS does not allow an individual to test the moral appropriateness of the decisions it derives from group datasets. Additionally, a data entry error in the system can immensely impact the defendant's future, and this can commonly occur when working with abundant cases or when inexperienced personnel use it. It can also decide between the different levels of long-term punishment. Hence, we are unable to comment on the algorithm's reasoning, even with control over its inputs and outputs.

With the use of automated decision making, consistency is not necessarily always the desired outcome and might even prove to be disadvantageous, particularly when target users are unfairly controlled by its use. AI systems can enforce laws and hence regulate the behaviour of targeted users by controlling how a decision is made and how users conduct their response to it.

Predictive tools in the legal system act without regard for demographic inequalities i.e. how frequently members of different racial, ethnic or religious groups are accused, arrested or convicted for various crimes. Hence, substituting a legal assistant for AI is not possible since,

while a judge might desire consistency to some extent, his primary aim should be to cater to the uniqueness and individuality of the case.

As far as I have researched upon RPS and SSS, and allocated time and effort in understanding their known application areas through authentic academic articles; they do not employ race, ethnicity, and religion as an explicit factor. There are various factors which can indirectly be linked to race, ethnicity, and religion in multiple plausible ways. For example, the algorithm might corelate country of origin, eye colour or educational background with the stated discriminatory factors. This is reasonable to some extent as, for example, the majority population of the defendant's country of origin might not be diverse and identify with a single race. Even in human reasoning, there is a marked difference between causation and correlation of factors. Regardless, explicit use of race, ethnicity or religion in AI is widely designated as inappropriate for the use cases of the CJS.

A risk assessment score does not necessarily become trustworthy and fair just because it was generated by AI, and the study from [4] is a testament to that. The study presents a fresh perspective to computerized risk assessment altogether. It found that COMPAS was no more accurate at predicting recidivism than a group of people with no experience in the legal domain. This is an alarming prospect indeed. These inexperienced candidates were provided with few features namely the defendant's age, sex and criminal history for a trial case. COMPAS's algorithm is only able to achieve the same level of prediction accuracy as a community of non-experts, and hence its credibility and fairness is severely undermined.

5. CONCLUSION

5.1 Conclusion

Governments need to regulate the applications of AI systems by expanding the powers of legal agencies to monitor this advanced technology. Applications of AI are expanding rapidly, without adequate governance or sufficient responsibility measures. The CJS, with its own regulatory framework and structural hazards, needs a sector-specific approach that focuses on the use of risk assessment tools and sentencing software.

Recidivism prediction and sentencing software are providing valuable assistance to the judiciary, regardless of the nature of the case. However, judges are increasingly relying on a computerized assessment that they cannot understand and explain, which is the most important reason to think twice before deploying it. The reasoning of the algorithm is not something that the court staff can personally verify. Therefore, the outcoming information is not necessarily accurate or applied in a meaningful way, especially in homicide cases. Hence, it is neither possible nor desirable ever that computers replace judges in the judicial function. If automated actuarial assessment cannot be made morally neutral or logically explainable, the CJS must dismiss it altogether, until it fits within relevant moral policies. Finally, it would be in the best interests of a convicted offender to get a proper schema of the decisions or sub-decisions which are affecting his life.

The use of risk assessment tools and sentencing software has initiated and enhanced several forms of criticism towards the CJS i.e. opacity in decisions, genuine concern that recidivism prediction and sentencing decisions are based on flawed datasets and perceived dehumanization of legal action. These criticisms, when mapped to homicide cases, can result in horrific and immoral outcomes. In choosing to employ predictive algorithms, the CJS should take advantage of global media to carefully address common challenges and convey the legal framework (if any) in which they operate. This ensures that the general public views its operations as rational and legitimate. This would also lead to increased support from both; the respective governments and the general public.

5.2 Reading Recommendations

Following is a list of the reading recommendations that can be held as grounds to extend the discussion on this topic:

[1] Hall, Maria & Calabro, Domenico & Sourdin, Tania & Stranieri, Andrew & Zeleznikow, John. (2005). Supporting discretionary decision-making with information technology: A case study in the criminal sentencing jurisdiction.

[2] Hutton, Neil & Patterson, A & Tata, Cyrus. (1995). Decision Support for Sentencing in a Common Law Jurisdiction. 89-95. 10.1145/222092.222145.

[3] Washington, Anne, How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate (February 4, 2019). Accepted for publication. The Colorado Technology Law Journal. Volume 17 Issue 1 http://ctlj.colorado.edu. Available at SSRN: https://ssrn.com/abstract=3357874

[4] Mayson, Sandra Gabriel, Bias In, Bias Out (September 28, 2018). 128 Yale Law Journal 2218 (2019), University of Georgia School of Law Legal Studies Research Paper No. 2018-35. Available at SSRN: https://ssrn.com/abstract=3257004

[5] Deeks, Ashley, Predicting Enemies (March 2018). Ashley Deeks, Predicting Enemies, 104 VA. L. REV. __ (2018, Forthcoming).; Virginia Public Law and Legal Theory Research Paper No. 2018-21. Available at SSRN: https://ssrn.com/abstract=3152385

6. BIBLIOGRAPHY

6.1 References

[1] Kehl, Danielle, Priscilla Guo, and Samuel Kessler. 2017. Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

[2] Fuchs, Daniel J. 2018. "The Dangers of Human-Like Bias in Machine Learning Algorithms." *Missouri S&T's Peer to Peer* 2, (1). https://scholarsmine.mst.edu/peer2peer/vol2/iss1/1.

[3] Michael Tonry, *Legal and Ethical Issues in the Prediction of Recidivism*, 26 Fed. Sentencing R. 167 (2014), *available at* http://scholarship.law.umn.edu/faculty_articles/525.

[4] J. Dressel, H. Farid, The accuracy, fairness, and limits of predicting recidivism. Sci. Adv. 4, eaao5580 (2018).

[5] Rubel, A., Castro, C., & Pham, A. (2018). Algorithms, Bias, and the Importance of Agency.

[6] Leah Wisser. (2019). Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing, *available at* https://www.law.georgetown.edu/american-criminal-law-review/wp-content/uploads/sites/15/2019/06/56-4-Pandoras-Algorithmic-Black-Box-The-Challenges-of-Using-Algorithmic-Risk-Assessments-in-Sentencing.pdf.

[7] UNODC (2014), Global study on homicide 2013: Trends, contexts, data, UNODC, Vienna, https://doi.org/10.18356/c1241a80-en.

[8] J. Angwin, J. Larson, S. Mattu, L. Kirchner, "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks," ProPublica, 23 May 2016; https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing