

TÓTH, SZIMONETTA

The Relationship Between AI Risk Analysis and Prosecution in the United States of America

Introduction

Many people believe that only humans are capable of bias, but machines are not. Humans create programs, which implies that, just as infants learn certain behaviors from their parents, computers are influenced by the algorithms and data sets that they are fed. Pre-existing disparities may emerge at any time, and they can be reproduced, which “Big Data” companies do not assist with – rather, these discrepancies are magnified and developed many times over. These are referred to as algorithmic biases.¹

Artificial intelligence is propelled forward by Big Data. Big Data and machine learning make it possible to automate human decision-making. These automations can occur in governance, criminal justice, or other sectors where previous injustices may be replicated as a result of algorithmic biases.

Risk analysis algorithms are employed to help decision-making in many sectors of life in today’s digital society, including criminal justice. The core premise of such risk assessment software is that it makes predictions on behalf of the predictive justice system, basing them on an established profile and employing probability calculations. Several types of crime forecasting can be distinguished based on their field of application: those that assist the police in their crime prevention work and law enforcement, and those that are sentencing-focused which are employed strictly in the field of criminal justice.

In the United States of America, the need and justification for the use of algorithm-based risk analyses has been driven by one of the biggest problems in the wider criminal justice system, which is overcrowding in prisons. The country has an extremely high prison population due to the strict use of criminal law as a deterrent for committing crime. This

¹ CARTER, Michael: *A technológia veszélyei a kisebbségek és az öslakosok jogaira nézve*. In: Grant, Peter (szerk.): Kisebbségek és öslakos népek 2020. Középpontban a technológia. Minority Rights Group International, London, 2020. 29. https://minorityrights.org/wp-content/uploads/2021/02/2020_MR_Report_Hungarian_Final_WEB.pdf (20.04.2022).

is, of course, far too costly. So, it is not surprising that in recent years economic austerity has become a central issue in the US criminal justice system. To prevent the overcrowding of prisons, all member states use various assessment tools based on computerized risk analysis.

Risk Analysis

Pre-sentence investigation reports (PSIs) (report prepared by a probation officer with the department of corrections in anticipation of a sentencing for a person who's either entered a plea to a criminal felony or is being sentenced after trial) have been used in the United States for decades, and in recent years a new element has been added to the information base in several member states: the information system-based risk analysis data.

The essence and basis of risk assessment software is that, on the basis of an established profile, probability calculations can be used to make predictions. Such risk assessment systems are also used by the police and crime prevention to determine which places become crime hotspots, or which people with a certain profile are more likely to commit a crime. It is important to link databases along data such as the persons involved, the places where the crime was committed, the means of offence... An artificial intelligence program operates this risk assessment software, the main purpose of which is to use probability to make predictions for predictive justice based on an established profile. The decisions made by AI and the risk assessments it performs also influence the chances of bail, conviction, and parole.² Risk analysis is also employed in the prison system as a technique for predictive forecasting. Algorithms are used to categorize convicts into low, medium, or high-risk categories based on pre-defined risk criteria, and to relate them to things like parole eligibility.³ In the case of sentencing, it is a matter of using mathematical methods to predict not only whether the accused is expected to appear at trial, but also what the ideal sentence would be, in terms of gender, level, and content, based on the defendant's criminal history and the type of offence committed, and which would serve both general and specific prevention purposes.⁴

The use of Big Data sets in predictive policing could logically lead to the use of Big Data sets for predicting prison sentences – an equally troubling concern, given the potential for algorithmic bias in the wake of a criminal justice system fueled by institutional racism.

In almost every state in the United States, AI-assisted pre-trial risk analysis is used before trials to assess the likelihood that a defendant will, for example, recidivate or even appear at trial.

² FANTOLY, Zsanett – LICHTENSTEIN, András: Számítógépes kockázatelemzés és büntetőeljárás. *Belügyi Szemle* 66. évf. 10. szám (2018) 5.

³ MONAHAN, John – SKEEM, Jennifer L.: Risk Assessment in Criminal Sentencing. *Annual Review of Criminal Psychology*, 12 (2016) 500. <https://www.annualreviews.org/doi/10.1146/annurev-clinpsy-021815-092945> (28.03.2022).

⁴ FANTOLY–LICHENSTEIN, 2018. 5–6.

Criminal statistics data collection can be divided into several generations. Following Cesare Lombroso's theory of criminal anthropology⁵, the first registers were produced in the early 20th century that summarized the non-physical variables that inclined people to commit crimes and were anticipated to detect potential offenders, the first of statistical risk analysis. The second generation supplemented the previous generation with characteristics of the criminal and the crime committed. The third generation examines dynamic factors; the fourth generation considers not only the offender's behavior and conduct, but also, for example, directly recommending a plan of action and treatment to the court and the prison service or advertising themselves as being able to predict whether the offender will appear in court.⁶

As it was already mentioned, in the United States of America AI assisted risk analysis is used in almost all member states; the assessment tools are the following: COMPAS⁷, LSI-R⁸, PSA⁹.

COMPAS

COMPAS is one of the most well-known of these fourth generation „automated decision support” risk analysis systems. COMPAS is a closed-source software developed by Northpointe. Northpointe's algorithm is protected by a license thus is not open to independent verification. This is also known as black-box theory¹⁰, since a black-box in systems theory typically refers to a device whose internal workings are not known, and therefore can only be examined by its input and output. The lack of transparency and the relative opacity of the operation of the algorithm and the evaluation of the data have opened the door to criticism. During hearings, judges are often shown the results of COMPAS, but defendants and their lawyers are not allowed to see it in its entirety.

Northpointe, the creator, claims that their system, which considers 137 criteria, can be used to forecast the defendant's recidivism with such accuracy probability¹¹ that it is currently being used in criminal procedures in various US Member States, such as punishment and sentencing. The risk factors can be many, but essentially the offender's

⁵ BROOKES, Elisabeth: Cesare Lombroso: Theory of crime, criminal man, and atavism. *Simply Psychology*, July 20 (2021) <https://www.simplypsychology.org/lombroso-theory-of-crime-criminal-man-and-atavism.html> 02.04.2022).

⁶ FANTOLY-LICHENSTEIN, 2018. 8–10.

⁷ Correctional Offender Management Profiling for Alternative Sanctions.

⁸ Level of Service Inventory Revised.

⁹ Public Safety Assessment.

¹⁰ BUNGE, Mario: A General Black Box Theory. *Philosophy of Science*, 30. évf. 4 (1963) 346–258. <https://www.jstor.org/stable/186066> (23.03.2022).

¹¹ BRENNAN, Tim et al: A Response to “Assessment of Evidence on the Quality of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS).” *Northpointe Institute for Public Management, Inc.* 2009. https://www.academia.edu/54493399/A_Response_to_Assessment_of_Evidence_on_the_Quality_of_the_Correctional_Offender_Management_Profiling_for_Alternative_Sanctions_COMPAS (25.01.2022).

personal characteristics of the offender, such as: age, gender, geographical environment, family background, employment or unemployment, economic situation, circle of friends, education, mental state, etc. Among these 137 criteria, five are the most essential¹²: (1) the offender's (deviant) lifestyle, (2) personal relationships/lifestyle, (3) offender's personality, (4) offender's family circumstances, (5) social relationships. In practice, it is used in parole hearings and in the area of sentencing.

COMPAS is the most widely used risk analysis system in the United States; its conclusions are widely acknowledged by experts, and its dependability is undisputed. COMPAS was designed to assist prison officials and probation officers in determining which therapy and re-socialization approaches would be most beneficial for certain offenders. But when employed in sentencing¹³, it can have a detrimental influence on the outcome of plea-bargaining cases since the results of COMPAS frequently encourage judges to contemplate imposing a more severe sentence on the offender than the one agreed upon by the clients during the plea negotiating process.

In 2016, ProPublica published a study on COMPAS¹⁴ and its results, calling the system remarkably unreliable. Programs are developed by humans – which means that just as children can learn the ethnic, religious or gender-based prejudices of their parents or community, these biases can seep into the machines and into their algorithms and data sets. The study described the flaw in the COMPAS system: non-recidivist black defendants received twice the recidivism rate of white defendants. In the case of a repeat offence, white defendants were considered about 50 per cent less likely to re-offend, meaning that when the algorithm thought black defendants were more likely to re-offend (compared to white defendants).

The developer of COMPAS, Northpointe, did not shy away from discussions with ProPublica¹⁵, sensing the fact that there was growing opposition to the scheme, both in jurisprudence and in the legal profession. After the criticism, COMPAS has tried to improve its system by introducing the "fairness criterion."¹⁶ If an algorithm has been corrected for bias against one group, this does not necessarily mean that it will not fail for other groups.

Some IT experts have also pointed out that¹⁷, although the use of algorithms seems to increase the efficiency of the process and the soundness of decisions, it can raise

¹² FANTOLY–LICHENSTEIN, 2018. 15–16.

¹³ Practitioner's Guide to COMPAS Core. Northpointe, 2015. <https://s3.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf> (22.01.2022).

¹⁴ ANGWIN, Julia et al: Machine Bias. *ProPublica*, May 23 (2016). <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (13.04.2022).

¹⁵ It should also be noted that after the publication of the study, Northpointe changed its name to Equivant.

¹⁶ FANTOLY–LICHENSTEIN, 2018. 10.

¹⁷ CORBETT-DAVIES, Sam et al: A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *The Washington Post*, October 17 (2016) <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublica/> (22.02.2022).

serious ethical and scientific problems. Their use should be constantly monitored, and in principle, a critical approach should be adopted to the possibility of these algorithms playing an increasingly significant role in the criminal justice system.

Loomis Case

The Loomis case is the most well-known COMPAS case (2013).¹⁸ The circumstances of the Loomis case in 2013 were Eric Loomis's prosecution for firing from a car and other lesser offenses. Mr. Loomis denied culpability at trial, admitting only that he was driving the car used in the crime considerably later than the time of the crime, but on the same night. During the sentencing process, the Wisconsin probation officer handed to the court a PSI database that was mostly compiled from the COMPAS risk analysis algorithm. Loomis was sentenced to six years' imprisonment and five years of enhanced criminal supervision by the court, taking into account the COMPAS results. Loomis appealed against this decision, arguing that the sentencing based on the COMPAS results violated his right to a fair trial. For instance, COMPAS does not explain how risk scores are established. Due to this lack of openness, defendants are unable to challenge the scientific validity and accuracy of such scores. Second, COMPAS reflects racial and gender prejudices. All else being equal, black men are more likely to be incorrectly projected to reoffend, whereas females are assigned a lower risk score. Third, the system bases its predictions on statistical correlations. Thus, the use of COMPAS by the court violates both the right to a personalized punishment and the right to be punished based on correct facts. The motion highlighted that in using the COMPAS results, the court had unconstitutionally relied on data, some of which allowed for racial discrimination, which was part of the COMPAS question base. Loomis's right to receive an appropriate, individualized sentence was thereby violated. Other grounds included that the sentence was not based on information from a specific knowledge base, as the court was not aware of the algorithm used to impose the sentence.

The Wisconsin Supreme Court judge took the legal position that the ruling was not unlawful. However, the decision emphasizes that the trial court did not consider only the COMPAS results, but also other factors in the sentencing judge's discretion. However, the judge also expressed personal concerns about the COMPAS system.

Expressed Concerns About the Use of COMPAS¹⁹

The system alone cannot be used for decision-making, it can only be one factor in the evidence. To summarise its advantages, it can be most useful where (1) it can be used to prevent a low-risk offender from being sent to a correctional institution; (2) it can be

¹⁸ State v. Loomis, 881 N.W.2d 759 (Wis. 2016)

¹⁹ KEHL, Danielle – GUO, Priscilla – KESSLER, Samuel: *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. Responsive Communities Initiative. Berkman Klein Centre for Internet & Society, Harvard Law School, 2017, 18-20. https://dash.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf (22.04.2022).

a factor in public safety by providing continuous supervision of dangerous offenders, even if this is not carried out in a correctional setting; (3) it can also help in the choice of probation, supervision and treatment models. The court also set age limits for the use of risk analysis systems. While it considers that these methods are undoubtedly useful in revealing the offender's motivation and personal circumstances, they should not be used exclusively to decide the level and duration of a specific sentence; nor should the result of the analysis alone be used as an aggravating or mitigating circumstance in sentencing. The main reason for this is that COMPAS does not consider all the relevant aspects of sentencing, but mostly examines certain factors through the recidivism perspective. Other sentencing circumstances (e.g., culpability, blameworthiness, deterrence) are not reflected in the algorithm. The Court would therefore make it mandatory for the judgment to make clear which aspects of the assessment were based exclusively on the COMPAS results, for example it would also impose a detailed obligation to state reasons in this respect.

Five Warnings from the Court for the Future Users of the COMPAS System

The application of COMPAS would therefore be limited to deciding matters concerning arrests under the Wisconsin Supreme Court's decision. The application of the risk analysis method is precluded from determining whether an offender should be sentenced to imprisonment, nor can it determine the severity of the sentence to be imposed. In addition to the prohibitions on its use, the court also issued five warnings to judges who intend to use the algorithm in the future.

- I. COMPAS is a proprietary tool, which has prevented the disclosure of specific information about the weights of the factors or how risk scores are calculated.
- II. and therefore identify groups with characteristics that make them high-risk offenders, not particular high-risk individuals.
- III. Several studies have suggested the COMPAS algorithm may be biased in how it classifies minority offenders.
- IV. COMPAS compares defendants to a national sample but has not completed a cross-validation study for a Wisconsin population, and tools like this must be constantly monitored and updated for accuracy as populations change.
- V. COMPAS was not originally developed for use at sentencing.²⁰

Judge Abrahamson's Dissenting Opinion

Abrahamson's opinion suggested that the developer of COMPAS, Northpointe, should provide information on the essence of the system's operation. She argued that this would rule out any suggestion that the algorithm takes too much account of the geographic location of the offence (or the offender) or the socio-economic status of the offender. The clarification should cover the following aspects: (1) what data (input factors) the

²⁰ Loomis, 881 N.W.2d 769–770.

algorithm uses; (2) how the algorithm weights each input factor; (3) which of these factors (e.g., race, gender) might be problematic in the evaluation formula.²¹

Algorithmic Accountability Act as a Solution

The Algorithmic Accountability Act of 2022 was introduced in the U.S. Senate as S. 3572²² and in the U.S. House of Representatives as HR 6580²³ on February 3, 2022. The bill proposes to direct the Federal Trade Commission (FTC) to promulgate regulations that require any “covered entity” to perform impact assessments and meet other requirements regarding automated decision-making processes and in particular those that implicate an “augmented critical decision process” – essentially, that result in any legal or other material effects – on a consumer. The Algorithmic Accountability Act would apply to companies that make over \$50 million per year, hold information on at least one million people or devices, or primarily act as data brokers that buy and sell consumer data.²⁴ These companies would have to evaluate a broad range of algorithms – including anything that affects consumers’ legal rights, attempts to predict, and analyze their behavior, involves copious amounts of sensitive data, or “systematically monitors a large, publicly accessible physical place.” That would theoretically cover a huge swath of the tech economy, and if a report turns up major risks of discrimination, privacy problems, or other issues, the company is supposed to address them within a timely manner.

An „augmented critical decision process,” as defined in the bill, is a process, operation, or other activity that employs automated decision systems to make a key choice. In contrast, an „automated decision system” is any system, program, or process (including those generated from machine learning and artificial intelligence approaches) that employs computing and the output of which serves as the foundation for a decision or judgment. The bill includes provisions for a „critical decision,” which is defined as any decision or judgment that has any legal, material, or similarly significant effect on a consumer’s life relating to or the cost, terms, or availability of, among other things, legal services, or any other service, program, or opportunity decisions that have a comparably legal, material, or similarly significant effect on a consumer’s life.

Conclusion

Computers are increasingly being used to make the most crucial choices in Americans’ life, such as whether they can purchase a home, get a job, or even go to jail. However, rather

²¹ Id. 774–75.

²² S.3572 – Algorithmic Accountability Act of 2022. <https://www.congress.gov/bill/117th-congress/senate-bill/3572/text?r=1&s=1> (03.05.2022).

²³ H.R.6580 – Algorithmic Accountability Act of 2022. <https://www.congress.gov/bill/117th-congress/house-bill/6580/text?r=5&s=1> (03.05.2022).

²⁴ L. BRYAN, Kristin – FATH, Kyle R. – TOMIMBANG, Gicel: Federal Lawmakers in House and Senate Introduce Algorithmic Accountability Act of 2022. *The National Law Review*, February 11 (2022) <https://www.natlawreview.com/article/federal-lawmakers-house-and-senate-introduce-algorithmic-accountability-act-2022> (28.04.2022).

than eradicating bias, these algorithms frequently rely on biased assumptions within the data itself, which can actively promote discrimination against women and people of color.

In the 21st century, technology has evolved so rapidly that we initially feared it, but as time went on, we discovered its potential. Based on the information already mentioned, artificial intelligence will inevitably be a part of the future, but it needs to be monitored. The involvement of minorities and indigenous peoples in the planning, implementation and evaluation of these programs could be a solution to the problems between technology and minority groups. Another remedy to these concerns could be to establish clear accountability guidelines for all AI, machine algorithm or other technology-enabled decision making to uphold the rule of law. Most importantly, and something that must be kept in mind for the future, is that technology will not be able to address the underlying social injustices, which will require human beings to solve.