Rohit Musti

AI & Sentencing

Introduction

AI is permeating our legal system at several levels. I will explore the scope and effectiveness of current AI implementations for sentencing. Then, I will examine some incompatibilities between AI and sentencing. Finally, I will combine these findings and offer my insights into how we can use AI effectively, without sacrificing our moral obligation of due process, as well as what, if any, policies we ought to adopt to formalize these checks and balances.

Incomplete Data in Risk Assessment

One prominent use case of AI in our court system is through "risk assessment" tools. These are algorithms that take a defendant's history into account, amongst other factors, and turn these qualitative data points into concrete numbers that predicts what level of risk they represent to the public. The risk factor can be calculated for a variety of purposes: from the likelihood they will commit another crime to the likelihood that the defendant will appear for their next court date.

This use case, however, has an inherent incompatibility. The most significant incompatibility of AI with the legal system for sentencing recommendations is incomplete data. This was most clearly demonstrated in the case of James Rivelli. James Rivelli was arrested for shoplifting, with a record of aggravated assault, multiple thefts and felony drug trafficking, yet only received a risk score of 3 out of 10. Despite these clear signs of risk, the algorithm classified him as a low risk defendant. It was wrong. Less than a year after this classification, he was charged with two felony counts of shoplifting. What went wrong? The algorithm that was used to determine risk didn't have access to the information from his Massachusetts record, where he had committed most of his prior crimes. This example demonstrates a major flaw with using machine learning algorithms. Without complete data, they are not useful. Humans are susceptible to these same flaws. However, a human can correct these flaws much more easily than a machine can. There are complicated mechanisms that need to be navigated to access these important data points. These often require signatures from judges and physically parsing through paper records. Given enough advancements in machine learning and automation, this will eventually be possible for a machine to do as effectively as a human, however, it is currently much easier for a human to gain access to the information, and process it. Furthermore, the systems and mechanisms for accessing information is not built for automation.

Supporters of these systems say that humans are susceptible to these same flaws. They contend that many times, judges and jurors act with incomplete information and that results in unjust rulings. My response to this claim is that humans are less susceptible to these types of errors. For example, a human prosecutor attempting to understand the risk of a client with a known criminal record would know to check the records of states and counties the criminal had lived in/committed the crimes in. While machine algorithms have proven more adept an analyzing data, we have not developed algorithms that tell us when we are missing data and where that data might be found. Until AI becomes more robust, we cannot, in good conscience, built flawed algorithms without significant human oversight in their use. As it stands, there is too much blind faith placed in this automation of analysis, as we saw in Rivelli’s case.

We must develop standards on what data is allowed to be collected by AIs, what data these AIs are allowed to be trained on, mechanisms for guaranteeing complete access to data, how much data is required, and how the risk assessment scores can be used. It is important that we agree upon what data is allowed to be used to train these algorithms. Otherwise, they could collect every conceivable data point possible about our actions. This could result in them using metrics that, while predictive, we find morally wrong to use in making determinations. We must also agree upon the minimum required data these algorithms are allowed to be trained on. This involves requiring a minimum number of data points and a diversity of data sources. Already, there are many algorithms that are being trained on biased data that become biased due to the data they are trained on. By increasing the diversity of these sources and the amount of data, we can avoid overfitting a model to a biased data set. It is also necessary to create mechanisms for accessing complete data on an individual. As it stands, it is extremely difficult to guarantee that all data about a person, relevant for rendering a decision, has been gathered and fed into a classifier. Without complete data, we will see many more cases like that of Rivelli. We must also ensure that the risk assessment scores used by these algorithms are applied properly. They should only be used to determine risk of fleeing if given bail, not in the sentencing of the criminals. These scores are not designed for these purposes and can result in dangerous decisions if misapplied.

Another helpful sentencing regulation is that we should start building mechanisms to make it easier to algorithms to access data. These could range from digitization to automation of signing off on permissions. It is important to start increasing the digital access points of our legal system so that we are prepared for a future where these algorithms play a more central role. Right now, we are limited in what information can be accessed. Even when information can be accessed, there aren't interoperable digital standards between state governments or between state and federal governments. By introducing standards as a part of this digitization movement, it will be easier to create algorithms to take advantage of them and we could guarantee that algorithms make decisions with complete information.

Rohit Musti

Un-Interrogability of Risk Assessment Algorithms

In the case of *Wisconsin vs. Loomis* there were sinister consequences of using these risk assessment tools. When the defendant was being processed, he was asked a series of questions that were eventually fed into a risk assessment tool called “Compas”. Fast-forward to the date of the trial. At this point, the judge had heard all the evidence of the trial and based on the evidence from the trial and the score from the risk assessment tool, the judge gave Loomis a long sentence. Loomis challenged this sentence on the grounds that he was not able to assess how the algorithm arrived at its risk assessment score, a key piece of uncontested evidence. This is extremely worrying. The algorithm in Compas assigned different weights to each answer the defendant provided given the moral values that it was trained on; these moral values are based on the the values that guided the judgements in the data. Thus, variations in these weightings, variations in the moral philosophy of the programmers or the data, impact the final “objective” number produced by the algorithm. By not being able to see the algorithm, there is no possible way for anyone involved with the trial to understand what weights the algorithm used to deliver its recommendation. What makes this even more problematic, is that even if the algorithm was allowed to be examined, it is likely to have been made, in part, using machine learning. A process that results in algorithms and outputs that we do not fully understand.

The state supreme court eventually ruled against Loomis, finding that knowledge of the outputs was a sufficient level of transparency. I find this unsatisfactory. The court’s reasoning is that all the information used by Compas was either publicly available or provided by the defendant himself. Because Loomis had every opportunity to verify whether or not the information was accurate, they believe that all the data it used was accurate. I believe that the true problem with using this black box approach to algorithms is that there is no opportunity to challenge the logic of the machine to reduce a sentence. Additionally, the court failed to recognize that the score was misapplied in this case. It is not appropriate to apply a risk assessment score in sentencing as it is a misapplication.

This case is emblematic of a larger flaw in using modern day machine learning algorithms to provide recommendations. We do not fully understand how the algorithms make their predictions. Until we can break the communication barrier between us and algorithms, we will not be able to fully understand their decision-making process. This makes it nearly impossible for a judge to use the output of these algorithms without, unknowingly, being influenced any potential bias in them.

Some would argue that it doesn't matter if we understand the predictions, as long as they are accurate. I find this notion problematic as I believe it denies due process. Due process, is referenced by the fifth amendment as, "*No person shall… be deprived of life, liberty, or property, without due process of law*". Due process is referenced by the fourteenth amendment as, "*nor shall any state deprive any person of life, liberty, or property, without due process of law*". These two references make it clear that due process is extremely important in passing judgement on any legal proceeding. The working legal definition of due process is flexible and whether or not it is violated is usually determined on a case-by-case basis. I believe that denying the opportunity to inspect how much weight the algorithm places on various components is a violation of due process. Due process is about fair and consistent applications of procedure. Without examining the algorithm, we cannot determine if it is internally or structurally biased.

The potential for bias comes from the data it is trained on. All machine learning trained tools rely on historical data. Because US history is so fraught with bias, the historical data is also full of this bias. Any algorithm trained on biased data will produce biased results. This was demonstrated in a study by ProPublica. The study concluded that black defendants were more likely to be wrongly labeled as high risk by algorithms that white defendants. A common response to this objection is that humans are inherently biased: what is wrong with using algorithms if they are as biased as humans? The problem is that humans are constantly trying to look for signs of bias and trying to eliminate them. Algorithms do not have this capacity. They only operate on data and adjusting the algorithms by hand after they've been created would result in unpredictable changes. Although human judges are biased, we shouldn't give them biased tools/mechanics to increase their bias.

One policy we ought to implement is a mechanism for reviewing the algorithms. This could take the form of mandating that any AI source code become open sourced or requiring that the source code is made available to a regulatory review board established by a group of countries. In order to make this an agreed upon global standard, each country must be able to put forth their own representative. This regulatory board would determine whether or not the appropriate data was used to train them and whether or not the algorithms are unbiased and adhere to basic human rights and standards. There are dangers to a board of members appointed by elected officials as it could easily fall victim to populism, as so many institutions have. Therefore, it is important to either establish requirements for appointees or severely limit their powers of review over the initial created regulations.

Broadness of sentencing

Another example of how AI is incompatible with our current legal system is the broadness of sentencing. In most situations, the judge and jury have significant leeway in deciding what sentence ought to be delivered. This sentencing process is extremely contentious and often balances the intent of the law with the letter of the law. Each individual judge has their own interpretation of how much leeway they are allowed to have when attempting to understand the intent of the law. The late Associate Justice Antonin Scalia was known for his focus on primarily interpreting the letter of the law. This is known as analytic jurisprudence. Another type of jurisprudence is sociological jurisprudence: stressing the social context and effects of laws. Just Sonya Sotomayor is known for her sociological jurisprudence.

Because of the very approaches to jurisprudence, it is very unlikely that two justices are presented with the same case and render the same ruling. As a society, this is important. As our values evolve, it is important that our justices reflect the will of the electorate. Although many celebrate Chief Justice John Marshall's foresight and brilliant legal argument laid out in *Marbury vs. Madison*, most people wouldn't want him on the bench today because he owned slaves. Machine learning algorithms do not evolve with our society and, thus, cannot arrive at sentences.

Any sentencing algorithm must be based on the most effective decisions from the past. That is to say that we shouldn't use all sentencings from history, rather we should use only the sentences that were shown to rehabilitate defendants. This will ensure that we don't amplify existing biases and that we are working in the interests of society. We can borrow an idea from blockchain to ensure that we have this work. We could establish a consensus requirement/mechanism that would allow justices/members of society to determine what decisions ought to be weighed the most. Through the current practice of deferring to higher courts and the most recent decision, this mechanism already exists in a form. I advocate for the formalization of this mechanism so that it can be systematically applied and ensure that any algorithm training on these data is being trained on cases that we believe are important as a society. If we wanted to use members of society to determine this, we could allow people to give their votes to others. For example, if a CS professor trusts a law professor more than themselves to decide what cases most closely reflect our society, then they can temporarily assign their vote to be cast in accordance with the law professor. This would allow us to have the closest form of direct democracy as possible while retaining the practicality of republican systems. Because aligning your vote with someone or taking it away could be as simple as a click, this system guarantees that no one's voice is misrepresented by an official who reneges on a promise.

Humans Judging Humans

Another incompatibility is the AI cannot fully replace a judge. Judges are interpreters of the law, appointed, ideally, for their legal acumen. Thus, as a society, we agree that our justices have a special role in safeguarding our laws. There is some sense that humans ought to be the ones interpreting the laws that directly affect us. Even if an AI could interpret a law more exactly, there is some sense that humans ought to be in charge of governing humans. This often referred to as the right to "self-determination". If we hand over control of our legal sentencing system to AIs, we have effectively surrendered this right to self-determination. This violates many of the principles of Western Society.

A maliciously trained algorithm, could easily produce biased, tyrannical recommendations. If we increasingly rely on these algorithms and do not develop mechanisms of explaining the decisions of these algorithms, then we have lost the ability for humans to be properly determining the punishment of other humans. We have surrendered that right to an Algorithm.

In the United States, there is an inherent right, and obligation, to rebel against a tyrannical government. The US Constitution and governing system is largely based on the writings of John Locke. His notions of natural rights to Life, Liberty, and Property, pervade our laws and government. In Locke's second treatise, he writes that humans have the right to "shake it off, freeing themselves from the usurpation or tyranny that the sword has brought down on them, until their rulers give them a form of government that they'll willingly consent to." This philosophy is echoed in the US Constitution: "when a long train of abuses and usurpations, pursuing invariably the same Object evinces a design to reduce them under absolute Despotism, it is their right, it is their duty, to throw off such Government". It is clear that a tyrannical AI algorithm would necessarily need to be cast aside. Unless these algorithms are constantly updated to reflect our evolving society, they could easily become engrained in our system and function as tyrannical relics of generations past.

I believe that an un-interrogable, machine-learning trained algorithm whose decisions are relied upon to determine the fate of a person within our justice system, would be tyrannical force. If there were no mechanisms to explain its decisions and allowances to rely upon it, then we would effectively be surrendering our autonomy to such a system. This a tyranical system that must be rebelled against in accordance with our government.

In order to prevent a scenario requiring rebellion against AI sentencing algorithms, I think that we also should put in policies that dictate how AIs can be used by justices in court. Specifically, we ought to ensure that the justices do not just blindly rule on cases based solely on the AIs interpretation of the fact. While AIs may be extremely useful in producing a descriptive analysis of the case, it must not offer prescriptive rulings. That power ought to be reserved for the justices who have been entrusted by citizens through elected officials or directly elected themselves. By placing strong limitations on how the output of these algorithms can be used, we can safeguard ourselves against tyranny of the past.

Conclusions

Rohit Musti

We need to create mechanisms and regulations to ensure that machine learning algorithms and AI systems do not amplify injustice in our legal system. These mechanisms must be regularly reviewed and updated, to reflect our evolving society. Through all of this innovation and application of systems, we must remember the humans at the center of these systems. It is important to ensure that we do not create a set of policies that would place us under tyranny. Rather, we should thoughtfully and intentionally design these levers so that human values and thought are driving the process, not faceless decisions that cannot be explained.

Rohit Musti

Sources

Jason Tashea. Courts Are Using AI To Sentence Criminals. That Must Stop Now. https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/

Vincent Southerland. With Ai And Criminal Justice, The Devil Is In The Data. **https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data**

Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

John Locke. Second Treatise of Government.

The United States Constitution.