Judging Our New Judges: Why We Must Remove Artificial Intelligence from Our Courtrooms Now

Kieran Duffy Newcomb

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/honors

Part of the Applied Ethics Commons, Applied Statistics Commons, Artificial Intelligence and Robotics Commons, Criminal Law Commons, Data Science Commons, Ethics and Political Philosophy Commons, Judges Commons, Legal Ethics and Professional Responsibility Commons, Philosophy of Science Commons, Probability Commons, and the Science and Technology Law Commons

Recommended Citation


This Senior Honors Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Honors Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact Scholarly.Communication@unh.edu.
Judging Our New Judges:

*Why We Must Remove Artificial Intelligence from Our Courtrooms Now*

Kieran Newcomb
University of New Hampshire
Senior Thesis
Professor Nick Smith
Spring 2024
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>5</td>
</tr>
<tr>
<td>Introduction</td>
<td>6</td>
</tr>
<tr>
<td>Key Word Definitions</td>
<td>10</td>
</tr>
<tr>
<td>Criminal Sentencing and Artificial Intelligence</td>
<td>15</td>
</tr>
<tr>
<td>Weapons of Math Destruction</td>
<td>26</td>
</tr>
<tr>
<td>Damage</td>
<td>28</td>
</tr>
<tr>
<td>Opacity</td>
<td>33</td>
</tr>
<tr>
<td>Scale</td>
<td>36</td>
</tr>
<tr>
<td>Crossroads</td>
<td>36</td>
</tr>
<tr>
<td>In a Capitalist Economy, Cost-Effectiveness Rules</td>
<td>37</td>
</tr>
<tr>
<td>We Must Remove This Technology from Our Courtrooms Now</td>
<td>44</td>
</tr>
<tr>
<td>Objections</td>
<td>45</td>
</tr>
<tr>
<td>Conclusion</td>
<td>59</td>
</tr>
<tr>
<td>Bibliography</td>
<td>62</td>
</tr>
</tbody>
</table>
Acknowledgements

There are so many people to thank that I could not possibly address them all here.

First, Professor Nick Smith for reminding me that it is a waste of life to spend time studying something you are not interested in. It turns out that following my passions has done me quite well. Oh, and also thank you for being my faculty advisor, my several-time URC advisor, my SURF advisor, my senior thesis advisor, and for guiding me throughout my law school application process.

Thank you to the entire faculty of the UNH philosophy department. I’m a better thinker because of you all, and—at the serious risk of making this far too corny—also a better person. My years spent in this department have been some of the best of my life.

Separately, I’d like to thank the late Professor Jennifer Armstrong. I wish I had told you how important your mentorship was to me.

A huge thank you to Abby Towers, Esther White, Paul Salek, Maddie Sarsfield, Kylee Rock, and Owen Miller—some of my best teachers. I cannot wait to see the great things you all do.

Finally, I’d like to thank my family, who have all listened to me talk about predictive algorithms and our criminal justice system ad nauseum. Thank you to: My endlessly curious grandpa, James Duffy, who has been one of my fiercest interlocutors (to everyone else’s irritation). My parents, Charles Newcomb, who is full of ideas, and Cathy Duffy, who challenges me to defend my ideas on the fly. Without you two, I would be nowhere (literally). My brother, Lachlan Newcomb, for being a paragon of cool-headedness. And, of course, my smart, beautiful, funny, soon-to-be wife, Lena Beth Schneider, who has provided me unwavering support from the moment I said “I think I might quit my job and go back to school” four years ago. I love you.
I can’t believe my luck that I’ve found myself surrounded by such amazing people.
Abstract

In this paper, I explore some of the ways in which artificial intelligence might enhance the sentencing process through recidivism prediction technology. Notably, this technology can increase the accuracy of risk predictions and the speed with which sentencing decisions are reached. I then show, however, that the recidivism prediction technology is likely to turn into what data scientist Cathy O’Neil calls a Weapon of Math Destruction. The potential harmfulness of this technology is due not to the inherent nature of the technology, but the symbiotic relationship it will have with our already harmful criminal justice system. I argue that the objective of implementing this technology is increased cost-effectiveness. It is against this metric that we will evaluate the technology’s success. Thus, if the technology makes our criminal justice system far more cost-effective—even if it proves to greatly increase harms done to society by the criminal justice system—we would be unlikely to substantially change our system once we have implemented it. Because of this, I argue that we ought to remove AI from our courtrooms now.
Introduction

“Software is eating the world,” declared software engineer Marc Andreessen in 2011.1 Now artificially intelligent algorithms are eating the software.2,3 From the predictive text in our smartphones, to the TV shows suggested to us, to the news stories we are shown, these algorithms are becoming an increasingly important determinant of the meaning of our lives.4,5,6,7 They influence where we work, what we eat, who we date, and how we vote.8,9,10,11,12 “Breaking AI News” headlines are becoming a daily occurrence as we try our best to keep up with the technology’s proliferation.13 With the passing of each day, artificial intelligence (AI) becomes a

greater feature of human life. The legal field is no exception to this mass integration.\textsuperscript{14,15,16} In the United States, judges already use AI-powered statistical programs to predict recidivism and issue prison sentences.\textsuperscript{17,18,19,20}

On the face of it, there is a great deal to be excited about regarding the way AI can enhance sentencing.\textsuperscript{21,22,23} Notably, it can increase the accuracy of risk prediction and the speed with which sentencing decisions are reached. This can help to ameliorate a number of system-wide issues—for example, the problematic amount of time impoverished people spend in jail awaiting trial.\textsuperscript{24,25,26} Thus, some argue that integrating this technology into our judges’ chambers

\begin{itemize}
\item \textsuperscript{14} Nikolsaia and Naumov, “Artificial Intelligence in Law,” 2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), October 6, 2020, https://doi.org/10.1109/fareastcon50210.2020.9271095.
\item \textsuperscript{16} Epps and Warren, “Artificial intelligence: now being deployed in the field of law,” Judges’ Journal, 59(1), 16-19.
\item \textsuperscript{17} Wisconsin Department of Corrections, “Compas,” https://doc.wi.gov/Pages/AboutDOC/COMPAS.aspx.
\item \textsuperscript{26} Nazish Dholakia, “How the United States Punishes People for Being Poor,” Vera Institute of Justice, September 21, 2023, https://www.vera.org/news/how-the-united-states-punishes-people-for-being-poor#:~:text=So%2C%20it’s%20no%20surprise%20that,a%20higher%20likelihood%20of%20conviction.
is a positive step forward, both improving the efficiency of the system and ameliorating some of the problems that plague it.27

However, there is also reason for serious concern.28,29,30,31,32 AI technology can become what data scientist Cathy O’Neil calls *Weapons of Math Destruction* (WMD). WMDs potential to cause harm is due not necessarily to its own inherent features, but to its symbiotic relationship with the system into which it is introduced.33 Because our criminal justice system already causes a great deal of harm, I argue that this technology will become one of these destructive forces.

Nonetheless, its promised cost-effectiveness makes it likely that our reliance upon it will only increase in the coming years. I argue that, over the next decade or so, AI is likely to replace human judges in *the role of determining prison sentences*. This does not mean that there will be no human judge—judges do much more than sentence.34,35 But these algorithms can amplify the harms already caused by the system, and their complexity and efficiency can prevent intervention. I think we should take this prospect quite seriously. In conclusion, I argue that we should remove this technology from courtrooms until we have made quite serious reforms to our criminal justice system.

---

Before moving forward, I must address what I will not address in this paper. There are a number of questions that undergird the following discussion of improving our system with AI: To what extent must we reform our justice system in order to safely implement AI? What is the purpose of our punishing offenders? What do we want our criminal justice system to accomplish? I will argue that we must remove AI from our courtrooms until we have made serious reforms to our criminal justice system—but say we do. Then it would be time for a discussion about criminal justice system reform, and I do not make much of an argument about what sort of reforms we should make, nor what the greater good should be for such reforms. The absence of this conversation from this paper may seem conspicuous, for it is a critically important topic. Nonetheless, I have left it out because I believe it will sidetrack us from what is a time-sensitive issue.

The debate about our purposes for punishment is highly contentious. People quickly and quite venomously disagree about why we punish criminal offenders (this tends to be a division that reflects political differences as well, often making the debate even more combative\textsuperscript{36,37,38,39}). Nonetheless, I have noticed that these same people seem to agree that we should not introduce AI into judicial deliberation—at least not right now. Thus, I circumnavigate this underlying topic so that we may better capitalize on the common ground between opposing parties in an effort to take what I believe is a necessary first step: remove the technology.


I have omitted these underlying issues not because I want to avoid them, but because I believe that if we do not remove the technology from our courtrooms, we will not be able to meaningfully have the debate at all. The system will effectively be sealed in place by the unfathomable efficiency offered by AI. I hope to avoid the problem most common to philosophy in which we sit around arguing over theory while changes are made around us. We must be careful to not get left behind. Once we have removed the technology from our judges’ chambers, we will indeed need to address these underlying question. I will address these topics in future papers, hopefully engaging this debate for the rest of my career.

Key Word Definitions

Some of the key terms I will use in this paper are: systems, models, artificial intelligence, machine learning, recidivism, sentencing, and plea bargains.

A system is a set of structures organized together as parts of a mechanism or interconnected network. We can understand our society as a system that is itself composed of many sub-systems, which are, in turn, composed of more sub-systems. Our justice system is one such sub-system, further composed of a lawmaking system, policing system, court system, prison system, and more. We aim to improve the systems we interact with through developing models, both formal and informal.

---

A model is “an abstract representation of some process … [that] takes what we know and uses it to predict responses in various situations.”\textsuperscript{45} Cathy O’Neil notes that everyone uses thousands of informal models to navigate the various systems in their lives.\textsuperscript{46} For example, we use informal models to guide our behavior in the workplace. We use information that we have gathered, both about specific colleagues and about the generally agreed upon standards of appropriate workplace behavior, to inform how to comport ourselves at work. The more information we gather, the stronger our model gets, and the better we get at navigating our workplace systems.

Similarly, we can create formal mathematical models to improve our efficiency in a system. A classic example of this is 1980s baseball teams hiring statisticians to analyze player data to optimize defensive strategies.\textsuperscript{47,48} This seriously improved defensive outcomes and thereby made the teams more competitive.\textsuperscript{49} The predictive algorithms I discuss in this paper are formal mathematical models that help better predict recidivism.

\textit{Artificial intelligence} is a broad term that refers to the ability of computers to process and perform tasks that mimic human intelligence.\textsuperscript{50,51,52} We should understand AI as having three calibers: artificial narrow intelligence (ANI), artificial general intelligence (AGI), and artificial superintelligence (ASI). At the writing of this paper, humanity has succeeded in creating ANI:

\begin{itemize}
  \item \textsuperscript{45} Ibid.
  \item \textsuperscript{46} Ibid.
  \item \textsuperscript{47} Ibid.
  \item \textsuperscript{49} O’Neil, “Weapons of Math Destruction.”
  \item \textsuperscript{50} Frederic I. Lederer, “Here there be dragons: The likely interaction of judges with the artificial intelligence ecosystem,” \textit{Judges’ Journal}, 59(1), 12-15 (2020).
\end{itemize}
technology that is specialized to do one specific task.\textsuperscript{53,54} These task-specific technologies present what appear to be different features of intelligence. Thus, for the various features of human intelligence, there exist corresponding programs.\textsuperscript{55,56,57} The purpose of the AI in something like ChatGPT is to mimic formal human written language patterns (these programs are called large language models\textsuperscript{58}), while the purpose of the AI present in recidivism prediction models is ideal rationality. Our technology remains “narrow” for now because each of these pieces of technology cannot accomplish the tasks of others—but as we bridge the gaps between each of these programs, we draw closer to creating AGI.\textsuperscript{59}

AGI is technology capable of many tasks, to the point that it is indistinguishable from human intelligence.\textsuperscript{60,61,62} However, if ChatGPT became capable of accomplishing the tasks of recidivism prediction models, we still would not necessarily call it AGI. This designation comes only when technology has amalgamated all, or nearly all, of the features of human intelligence, such that it is indistinguishable. Although the difference between the ANI that we have today and

\textsuperscript{57} McGill, Justin. “How Many AI Tools Are There?” Content @ Scale, June 27, 2023. https://contentatscale.ai/blog/how-many-ai-tools-are-there/.
\textsuperscript{59} Tim Urban, “The Artificial Intelligence Revolution: Part 1.”
\textsuperscript{60} Tim Urban, “The Artificial Intelligence Revolution: Part 1.”
\textsuperscript{61} IBM, “What Is Artificial Intelligence (AI)?”
AGI is considerable, we progress daily. As Wharton Professor Ethan Mollick has said, the AI that you use today is the worst AI you will use for the rest of your life—and this will be true every time you use it.

Leading AI philosopher Nick Bostrom writes that when humans achieve AGI, we will soon thereafter have superintelligence. This is because intelligence begets intelligence asymptotically: “[o]nce artificial intelligence reaches human level, there will be a positive feedback loop that will give the development a further boost. AIs would help constructing better AIs, which in turn would help building better AIs, and so forth.” Quite quickly, this will result in ASI, which Bostrom describes as “an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.”

People oftentimes conflate AI with ASI, but this is a level of technology that we have never seen,

---

68 Computer scientist Ray Kurzweil calls this the Law of Accelerating Return, that technological progress happens more quickly over time (Tim Urban, “The Artificial Intelligence Revolution: Part 1.”). He argues that “the progress of the entire 20th century would have been achieved in only 20 years at the rate of advancement in the year 2000—in other words, by 2000, the rate of progress was five times faster than the average rate of progress during the 20th century. He believes another 20th century’s worth of progress happened between 2000 and 2014 and that another 20th century’s worth of progress will happen by 2021, in only seven years. A couple decades later, he believes a 20th century’s worth of progress will happen multiple times in the same year, and even later, in less than one month. All in all, because of the Law of Accelerating Returns, Kurzweil believes that the 21st century will achieve 1,000 times the progress of the 20th century.” We can readily see this in AI programs created in the last 20 years (Henshall, “Why AI Progress Is Unlikely to Slow Down.”). One program called MNIST, created to recognize handwriting, began in 1998 and only surpassed human ability in 2013. A more recent program called SQuAD 1.1, developed to complete reading comprehension tasks, began in 2016 and only took until 2017 to surpass human ability.
69 Nick Bostrom, “How Long Before Superintelligence?”
70 Nick Bostrom, “How Long Before Superintelligence?”
and probably cannot properly imagine.\textsuperscript{71,72,73,74,75} It is important to note where humanity stands in this path towards superintelligence, because we must simultaneously prevent the very common misunderstanding that all AI is ASI and understand that ANI can be exceedingly powerful.

*Machine learning* refers to the feature of artificial intelligence that imitates human learning, gradually improving its accuracy without explicitly being programmed to do so.\textsuperscript{76,77,78} Predictive AI uses algorithms to predict future data points based on its analysis of data input. Not only is it capable of handling enormous data sets, but it “learns” from new experiences. This means that when the program receives new data, it modifies its own algorithm to accommodate the new information.\textsuperscript{79}

*Recidivism* refers to when an individual is released from prison and reoffends.\textsuperscript{80} This is a metric used by criminologists and judges to identify a pattern of troubling behavior over time. *Sentencing* refers to when a judge declares the punishment for a defendant who has been found guilty of a crime.\textsuperscript{81} There are a range of possible sentencing outcomes, including probation, imprisonment, fines, boot camps, rehab, and house arrest.\textsuperscript{82}

\textsuperscript{71} There is fairly serious disagreement about whether humanity is close to creating ASI. There is also quite serious concern about what this will look like. Some fear that it will lead to the almost instantaneous end of humanity.\textsuperscript{72} IBM, “What is Artificial Superintelligence?” IBM. Accessed April 6, 2024. https://www.ibm.com/topics/artificial-superintelligence#:~:text=Artificial%20superintelligence%20(ASI)%20is%20more%20advanced%20than%20any%20human.
\textsuperscript{76} Allen, “Artificial intelligence in our legal system.”
\textsuperscript{80} National Institute of Justice, “Recidivism,” https://nij.ojp.gov/topics/corrections/recidivism.
\textsuperscript{81} Cornell Legal Information Institute, “Sentencing,” https://www.law.cornell.edu/wex/sentencing.
Only a small number of cases arrive at a sentencing hearing through trial.\textsuperscript{83} The vast majority of criminal cases end with a \textit{plea bargain}, which is when a defendant pleads guilty to the crime they are being charged with in hopes of receiving a lighter sentence.\textsuperscript{84,85,86} Although plea bargains often include a sentence agreed upon by the prosecutor and the defendant—and these sentences are often accepted—judges may overrule the agreed upon sentence based on their own view of the case and the defendant.\textsuperscript{87,88}

Criminal Sentencing and Artificial Intelligence

Criminal sentencing is a delicate, complex process.\textsuperscript{89} Deliberation can take weeks to months.\textsuperscript{90} The judge receives input from a number of sources: mandatory minimums and maximums set by Congress, sentencing guidelines set by the U.S. Sentencing Commission, and pre-sentence reports with statements from the victims, defendant, and attorneys.\textsuperscript{91} In reviewing these pre-sentence reports, the judge considers mitigating factors, such as the defendant’s

\textsuperscript{88} Andrew van Dam, “Algorithms Were Supposed to Make Virginia Judges Fairer. What Happened Was Far More Complicated.”
\textsuperscript{90} Offices of the United States Attorneys, “Sentencing.”
\textsuperscript{91} Ibid.
criminal history, whether he expresses remorse, the nature of the crime, mental health history, and much more.\textsuperscript{92,93,94,95} Thus, the sentence produced is a combination of legal guidelines, highly trained legal reasoning, and subjective assessments of the defendant’s perceived risk.\textsuperscript{96}

The nature of this risk assessment is a matter of debate.\textsuperscript{97} Some argue that judges’ training enables them to employ a scientific, mechanical application of law.\textsuperscript{98} The mandatory minimums and maximums and the sentencing guidelines—themselves primitive models—act as guardrails, and any potential bias involved in the risk assessment is attenuated through rigorous training in legal reasoning.\textsuperscript{100} Yet, others argue that risk assessment amounts to a judgement of character, which is necessarily subjective and therefore includes some level of bias.\textsuperscript{101,102,103,104,105,106,107} The judge might consider questions like: Does the person in question \textit{seem} like the type to commit crimes? Do they come from a good family and neighborhood? Do

\begin{thebibliography}{100}
\bibitem{92} Ibid.
\bibitem{94} New York State Unified Court System, “Sentencing Basics.”
\bibitem{95} Cassia Spohn, “How Do Judges Decide?: The Search for Fairness and Justice in Punishment.”
\bibitem{98} Guthrie, Rachlinski, and Wistrich, "Blinking on the Bench: How Judges Decide Cases."
\bibitem{100} Ibid.
\bibitem{101} Guthrie, Rachlinski, and Wistrich, "Blinking on the Bench: How Judges Decide Cases.”
\bibitem{107} Shawn D. Bushway, “Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
\end{thebibliography}
they associate with people who commit crimes? Did they go to college? Do they express what the judge believes to be genuine remorse? The judge assesses risk based upon a number of data points that appear to really just be value judgments.

Predictions for recidivism have been used by judges to inform parole decisions for nearly a century, with (non-AI-powered) statistical analytic software making its first appearance about two decades ago. Indeed, when recidivism risk is a legally relevant factor, judges in many states are urged to consult prediction models. However, it should be noted that using predictions about the likelihood of future criminal behavior to inform punishment is inductive. Because of the severity of the task at hand, it seems morally problematic to use inductive reasoning to justify punishment. Common sense tells us that punishment should be associated with the crime committed, not the crime predicted. But it simultaneously seems obvious to do so in order to guard the safety of the public: judges should distinguish between the person who displays remorse about killing in self-defense and the person who killed out of interest and shows no remorse. The practical question at the heart of this distinction is whether

---

108 Given that some people who have been found guilty or who have pled guilty are innocent, this is a tricky consideration. If the defendant is being genuine, they should not appear remorseful. But this lack of remorse may also negatively impact the outcomes of their sentencing.  
115 Lin et al., “The Limits of Human Predictions of Recidivism.”  
offenders presents a continuing threat to those around them. So, while this may seem fundamentally questionable, it’s an important part of sentencing.

Recidivism prediction is a model that aims to make our criminal justice system better. The introduction of statistical programs replaced what was traditionally an informal risk analysis with a formal mathematical model. AI integration is an update to that formal model. There are two metrics of performance that may be enhanced by this technology: speed and accuracy. The processing speed of AI is likely quite apparent to anyone who has used any form of AI-power technology, so I will address this relatively quick point before moving on to the much more complicated matter of accuracy.

Human beings are not particularly efficient at reviewing case history and writing judicial decisions. Consider how long it takes to conduct research and write a substantial paper. It is simply a time-consuming process, requiring an appreciable amount of reading, analysis, and writing—even for the very best human experts. This is why judges have clerks: recent law school graduates who conduct legal research, prepare bench memos, draft orders and opinions, proofread the judge’s orders and opinions, verify citations, communicate with counsel regarding case management and procedural requirements, and assist the judge during courtroom proceedings.

120 Adam Neufeld, “In Defense of Risk-Assessment Tools.”
122 Ibid.
Artificial intelligence, on the other hand, is immensely skilled at analyzing massive datasets.\textsuperscript{124,125,126} With the predictive coding available through AI technology, a defendant fills out a questionnaire and the judge has an algorithm provide the statistical likelihood of recidivism, account for sentencing rules and guidelines, and produce a sentence suggestion. With only a very modest software update to include something like ChatGPT, it could then provide a written explanation of how it determined the sentence.

Rather than taking weeks or longer, this could take minutes.\textsuperscript{127,128} Such increased speed is of critical importance to a system currently bottlenecked with a huge backlog of cases.\textsuperscript{129,130,131,132} Many, presumed innocent, spend months in jail awaiting trial if they are unable to pay bail.\textsuperscript{133,134,135} Of course, there are very topical concerns right now over AI programs like

\textsuperscript{124}Lederer, “Here there be dragons: The likely interaction of judges with the artificial intelligence ecosystem.”
\textsuperscript{125}Bringsjord & Govindarajulu, “Artificial Intelligence.”
\textsuperscript{126}IBM, “What Is Artificial Intelligence (AI)?”
\textsuperscript{129}This is particularly true for immigration courts right now.
\textsuperscript{135}Nazish Dholakia, “How the United States Punishes People for Being Poor.”
ChatGPT fabricating case history, but this is a speedbump on the road to what would have been incomprehensible efficiency ten years ago.\textsuperscript{136,137,138}

Let us turn now to accuracy. As previously mentioned, several decades ago judges began to have statistical programming made available to inform this part of their deliberation.\textsuperscript{139} In older statistical programs, pairs of variables would be run against each other, analyzing binomial relationships and their associated recidivism rates to determine a bivariate risk threshold.\textsuperscript{140} For example, age and level of education might be used to create a binomial graph showing the range that indicates the greatest likelihood of recidivism. The judge might then create a number of different graphs with the data points she believes to be most relevant and use each of these to inform her decision. This would not provide something as definite as a risk score, but the judge might learn which variable relationships provide the most meaningful insight and weigh those graphs more heavily (an informal model at work).\textsuperscript{141}

As our technology improves, however, scientists develop predictive programs that can handle far more complex variable relationships.\textsuperscript{142,143} While an old program might provide an assortment of graphs with “thresholds of predicted criminality,” new programs are able to

\begin{thebibliography}{99}
\bibitem{138} Admittedly, it probably doesn’t feel that way to the lawyer facing sanctions for using ChatGPT to write his brief.
\bibitem{139} Ludwik and Mullainathan, “Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System.”
\bibitem{140} Berk and Bleich, “Statistical Procedures for Forecasting Criminal Behavior.”
\bibitem{141} Ibid.
\bibitem{142} Ibid.
\bibitem{143} Bringsjord and Govindarajulu, “Artificial Intelligence.”
\end{thebibliography}
provide a simple risk score.\textsuperscript{144,145,146,147} This is important statistically, because the relationship between two variables can change with the presence of a third variable.\textsuperscript{148} For example, imagine we want to predict whether someone will commit a crime in the future, and we have two variables from which we can make our prediction: A and B. We might then predict, based upon the relationship between those two variables and the criminal history of people with similar characteristics, that the individual is unlikely to commit a crime. However, if we incorporate a third variable, C, this prediction might change considerably. This is because two of a defendant’s data points do not interact in isolation; the defendant is the composite of all of his data points. Our behavioral patterns—and therefore predictions about our behavior—change, dramatically or very slightly, with each variable that is added to the equation. Having technology that can accommodate this great complexity is extremely helpful.

Nonetheless, it is still sometimes argued that humans remain superior at risk assessment. In Dressel and Farid’s 2018 study, human subjects and the COMPAS system\textsuperscript{149}—which stands for Correctional Offender Management Profiling for Alternative Sanctions—were each given information from twenty old court cases and asked to predict whether the defendant reoffended after their release from prison.\textsuperscript{150} COMPAS, which is already used in courtrooms across the

\textsuperscript{144} Berk and Bleich, “Statistical Procedures for Forecasting Criminal Behavior.”
\textsuperscript{145} O’Brien and Kang, “AI in the Court: When Algorithms Rule on Jail Time.”
\textsuperscript{146} Dressel and Farid, “The accuracy, fairness, and limits of predicting recidivism.”
\textsuperscript{149} COMPAS is at the center of a Wisconsin court case that has garnered significant attention over the past several years. Eric Loomis, charged with assault, was sentenced to prison by a judge who consulted COMPAS for Loomis’ recidivism risk assessment. Notably, COMPAS recommended a six year sentence, which the judge issued. The judge’s use of COMPAS in this case disturbed many people because Dressel & Farid’s study showed it to be only about as accurate at recidivism prediction as randomly selected non-experts.
\textsuperscript{150} Dressel and Farid, “The accuracy, fairness, and limits of predicting recidivism.”
country, predicts recidivism by weighing the answers to a 137-item questionnaire.\textsuperscript{151,152,153,154,155} The questionnaire asks about things like age, criminal history, level of education, gang affiliation, economic status, levels of boredom, and anger issues.\textsuperscript{156,157,158} The 800 participants had an average accuracy rate of 66.7\%, while COMPAS had an accuracy rate of 65\%.\textsuperscript{159} Dressel and Farid argued that an algorithm so simple it could be written on the back of a business card—essentially cross-analyzing age and number of prior convictions—yields a similar accuracy rate.\textsuperscript{160}

However, later studies have disputed Dressel and Farid’s findings. One study conducted by Lin et al., suggested that Dressel and Farid’s study elicited human responses in a manner that aided their performance—meaning that the human accuracy rate was higher than it would have been in a natural setting.\textsuperscript{161} In fact, Lin et al.’s study showed that COMPAS achieved an 89\% accuracy rate, while humans achieved 83\% when given feedback on each case, and only 60\% when given no feedback.\textsuperscript{162}

\textsuperscript{156} Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”
\textsuperscript{157} Ed Yong, “A Popular Algorithm Is No Better at Predicting Crimes than Random People.”
\textsuperscript{158} Lin et al., “The Limits of Human Predictions of Recidivism.”
\textsuperscript{159} Dressel and Farid, “The accuracy, fairness, and limits of predicting recidivism.”
\textsuperscript{160} Though the program has also been criticized for its complexity (Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”).
\textsuperscript{161} Lin et al., “The Limits of Human Predictions of Recidivism.”
\textsuperscript{162} Ibid.
One might still object that judges undergo rigorous legal training before becoming judges. Don’t judges improve in their ability to predict recidivism through all of their training? Probably, but it is hard to tell how or by how much.\textsuperscript{163,164,165} Let us grant, though, that the rigorous training involved in legal practice significantly improves a human’s ability to predict future criminal behavior. It probably does, so this is not a particularly generous concession.\textsuperscript{166} Say it improves human prediction accuracy to 90%—a 12.5\% increase over the participants in Lin et al’s study who received feedback on their prediction after every case, and about a 35\% increase over those in Dressel and Farid’s study. The issue is at what point in AI’s progression we consider it irresponsible to let the human continue to make recidivism predictions.\textsuperscript{167} Is it 95\%? At 95\% AI prediction accuracy, we would be sacrificing 5.5\% accuracy to allow a judge to make his risk prediction. If we know that the AI program is more accurate, would it not be irresponsible to let the judge continue to make the assessment? The data tells us what we need to know—and it does so without subjective value judgments.

Indeed, it is argued that regardless of how much rigorous legal training a judge has, it is impossible to sanitize this process of implicit bias.\textsuperscript{168,169,170,171,172} This is coupled with the fact that humans do not seem to be particularly good at predicting the future behavior of other

\textsuperscript{164} Lin et al., “The Limits of Human Predictions of Recidivism.”
\textsuperscript{165} Bushway, “Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
\textsuperscript{166} Lin et al., “The Limits of Human Predictions of Recidivism.”
\textsuperscript{168} Guthrie, Rachlinski, and Wistrich, “Blinking on the Bench: How Judges Decide Cases.”
\textsuperscript{169} Smith, “Disparity in Context: Judges’ Perspectives on Disparities in a Sentencing Guideline System.”
\textsuperscript{170} Lartey, “US Prison Sentences Could Vary by up to 63\% Depending on Judge – Study.”
\textsuperscript{171} Kenthirarajah et al., “Does ‘Jamal’ Receive a Harsher Sentence Than ‘James’? First-Name Bias in the Criminal Sentencing of Black Men.”
\textsuperscript{172} Bushway, “Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
humans, including recidivism. Mental models are inherently vague and unrefined—better described more like gut feelings than calculations. This may be why we value years of experience—more experience equips the judge with an expert sense. There is something intangible—perhaps spiritual—happening when we assess someone’s character; we oftentimes describe it as something you can “sense” by “looking into someone’s eyes.”

Such an impressionistic approach makes it difficult for a judge to describe with any sort of scientific accuracy how they have improved in their ability to predict risk. Even if Dressel and Farid were correct and COMPAS was only as accurate as randomly selected non-experts in 2018, algorithms like COMPAS are fed more data every day and will refine their algorithms continuously for as long as we allow them to. Humans do not receive feedback as continuously—nor do they incorporate the feedback they do receive into their models as effectively. Humans tend to modify their informal models until they seem good enough to serve them in whatever system they are navigating. Any valuable improvements made die with them, limiting the human judiciary’s potential to make continuous progress in its prediction

175 Neufeld, “In Defense of Risk-Assessment Tools.”
176 Lin et al., “The Limits of Human Predictions of Recidivism.”
182 Lin et al., “The Limits of Human Predictions of Recidivism.”
accuracy. A program like COMPAS does not face this barrier. It can continue to improve open-endedly—and it is only improving faster.\(^{184}\)

The practical implications of this increased accuracy may be better illustrated by looking to another sector that has been profoundly changed by AI: social media marketing.\(^{185}\) Social media advertisements have become so personalized that some claim they feel as if advertising companies can read their minds.\(^{186,187}\) People comment that they receive advertisements for products they have only privately thought about. This is not because ad agencies can read minds, but because user data from social media platforms tells these programs information about the individual that allows them to predict what they are likely to want. This can be done with such great accuracy that sometimes it *seems* like they can read minds.\(^{188}\)

Social media companies accumulate massive amounts of data about their users—where they live, the political make-up of their environment, their political allegiance, places they have “checked in,” the clothes of influencers they follow, the type of music they listen to, who they are friends with, and so much more—and sell that information to retail companies.\(^{189,190,191,192}\)

This enormous volume of data being fed to extremely powerful predictive models is what allows

---

\(^{184}\) Epps and Warren, “Artificial intelligence: now being deployed in the field of law.”

\(^{185}\) O’Neil, “Weapons of Math Destruction.”


\(^{188}\) O’Neil, “Weapons of Math Destruction.”


companies to accomplish such jarring personalization. This is the nature of our data. It tells a story that can be so complete as to feel like this machine is capable of reading our minds. What if we could predict who the dangerous criminals were with such accuracy that it was like we were reading their minds?

Although recidivism prediction software has not quite established itself as clearly superior, it is clear how valuable this would be. Human sentencing is imperfect—frequently criticized as inconsistent or arbitrary. Sentencing guidelines have been created and refined, only to still be criticized. Despite all of the guardrails in place and the refined legal reasoning, sometimes judges are wrong in their assessments of defendants.

Weapons of Math Destruction

It is clear why many see the value in upgrading our risk assessment procedures. AI offers humanity this upgrade. Despite these considerable benefits, though, there is reason for

---

194 In fact, the very language that I use to describe this potential accuracy itself tells us something about the type of mystical ability we assign to the ability to accurately predict someone’s behavior. Artificial intelligence offers to free us from this sort of archaisism. We no longer need to rely on training our officials to “read” someone’s body language in hopes that we can see into their soul. No need for metaphysics and character assessments—we can predict their behavior with scientific precision.
195 This is currently a matter of debate.
196 Dressel and Farid, “The accuracy, fairness, and limits of predicting recidivism.”
197 Smith, “Disparity in Context: Judges’ Perspectives on Disparities in a Sentencing Guideline System.”
199 Lartey, “US Prison Sentences Could Vary by up to 63% Depending On Judge.”
201 Lederer, “Here there be dragons: The likely interaction of judges with the artificial intelligence ecosystem.”
202 Lin et al., “The Limits of Human Predictions of Recidivism.”
205 Neufeld, “In Defense of Risk-Assessment Tools.”
206 Lin et al., “The Limits of Human Predictions of Recidivism.”
serious concern about AI-powered risk assessment. Cathy O’Neil calls these recidivism prediction models *Weapons of Math Destruction*: math-powered programs which encode human prejudice, misunderstanding, and bias into the software system.207 “Like gods,” she writes, “these mathematical models were opaque, their workings invisible to all but the highest priests in their domain: mathematicians and computer scientists. Their verdicts, even when wrong or harmful, were beyond dispute or appeal. And they tended to punish the poor and the oppressed in our society, while making the rich richer.”208

O’Neil identifies three hallmark features of WMDs—*damage, scale, and opacity*—which serve to make these software programs especially harmful.209 It is not the technology alone that is destructive: models work symbiotically with the system in which they are implemented, being shaped by the system and co-creating it. While AI integration will likely result in tremendously increased efficiency, if the system is harmful, it may very well increase the severity and frequency of these harms.210,211,212,213,214 Moreover, increased efficiency could make the system more difficult to meaningfully change later.215 Thus, using this technology could make our criminal justice system—already criticized as severely systemically biased—far more

---

208 Ibid.
209 Ibid.
210 Harcourt, “Risk as a Proxy for Race: The Dangers of Risk Assessment.”
215 Ibid.
Let us consider how the three hallmark features of O’Neil’s WMDs map onto the recidivism prediction technology used by judges.

**Damage**

Does the use of this technology in our courtroom have the ability to cause damage? Since this technology simply promises increased efficiency of the system already in place, the question should really ask whether our criminal justice system causes damage. In a way, our criminal justice system is fundamentally concerned with causing damage. Generally speaking, criminal justice systems punish people who have committed crimes.226 Juridical punishment is an “intentional infliction of harm on an individual for an offense against law by those vested with

---

223 Bushway, “‘Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
political authority to do so.”\textsuperscript{227} Therefore, it is somewhat obvious to claim that our criminal justice system causes harm—that is its purpose.\textsuperscript{228} This technology will make our justice system more efficient. So, yes, this technology has the ability to do damage, because that is the task at hand.

Perhaps the question we should really ask is whether this technology will produce unjust harm. Ideally, every time the criminal justice system harms someone, it is just—meaning, basically, that the individual has been judged to deserve it.\textsuperscript{229} However, there is evidence to suggest that our criminal justice system produces troubling amounts of unjust harm. This has a dual meaning. First, there are people imprisoned who have not earned it in a literal sense—a disturbingly high number of convicted people are later found to be innocent.\textsuperscript{230,231,232} This is a tragedy, which the United States is officially committed to avoiding\textsuperscript{233}—though the pressure with which defendants are urged to accept plea bargains suggests that this is more a verbal commitment than an ideal the system truly strives to achieve. Second, our criminal justice system produces harms outside the prison walls, often pulling people from systemically disadvantaged


\textsuperscript{228} Why punish at all, though? This is a contentious debate that, unfortunately, falls outside the scope of this paper. There are four commonly agreed upon aims for punishment: retribution, deterrence, incapacitation, rehabilitation, and restitution (Monahan & Skeem, “Risk Assessment in Criminal Sentencing”). It is sometimes claimed that the United States’ massive prison system accomplishes each of these aims, but most often we do not explicitly justify our massive prison system. There certainly are debates about the state of our prison system, but rarely its existence in the first place. Rather, many Americans have an assortment of implicit beliefs about why and how we punish, which we rarely question very critically.


\textsuperscript{230} Georgia Innocence Project, “Beneath the Statistics: The Structural and Systemic Causes of Our Wrongful Conviction Problem.”


communities into the prison system, and then keeping them coming back after they have been released.\textsuperscript{234,235,236,237,238} Objections to this second point abound—I will address some of them later in this paper.

The hope for this new technology is that it will reduce some of these harms by attenuating bias and producing more even-handed, scientifically-informed sentencing.\textsuperscript{239,240,241} An obstacle to this is that the technology learns by analyzing data from past court cases.\textsuperscript{242,243,244,245} Our criminal justice system has historically been mired in open discrimination—and many argue it still is.\textsuperscript{246,247,248,249,250} For example, Black Americans

\begin{footnotesize}
\begin{enumerate}
\item This is a hotly contested topic. There is disagreement in the United States about whether our criminal justice system, and other systems, are structured such that people—primarily poor and black or brown—are pulled into the criminal justice system. I make the argument here that this is the case, but this is a claim that is not without challenge.
\item Davis, “Are Prisons Obsolete?”
\item Reiman and Leighton, “The Rich Get Richer and the Poor Get Prison.”
\item Bushway, “Nothing Is More Opaque than Absolute Transparency”: The Use of Prior History to Guide Sentencing.”
\item Silberg and Manyika, “Tackling Bias in Artificial Intelligence (and in Humans).”
\item Heilweil, “Why Algorithms Can Be Racist and Sexist.”
\item Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis.”
\item Davis, “Are Prisons Obsolete?”
\item Reiman and Leighton, “The Rich Get Richer and the Poor Get Prison.”
\item O’Neil, “Weapons of Math Destruction.”
\end{enumerate}
\end{footnotesize}
comprise 38.7% of the prison population despite being only 13.6% of the country’s overall population.\textsuperscript{251,252} Black men also receive sentences 13.4% longer and Hispanic men receive sentences 11.2% longer than White men, while Hispanic women receive prison sentences 27.8% longer and all other non-White women receive sentences 10.0% longer than White women.\textsuperscript{253} The U.S. Sentencing Commission suggests that, rather than disparities in prison sentence length, this contrast is due mostly to differences in probationary sentences. Black and Hispanic men are shown to be 23.4% and 26.6% less likely, respectively, to be sentenced to probation than White men; Hispanic and Black women are 29.7% and 11.2% less likely, respectively, to receive a probationary sentence than White women. Additionally, impoverished people are much more likely to spend time in jail and be sentenced to prison.\textsuperscript{254,255,256} The median pre-incarceration salary is approximately $25,400, and nearly 80% of inmates have a pre-incarceration salary of less than $40,000.\textsuperscript{257,258} This is understood to be a result of poor peoples’ inability to pay fees and fines, the criminality of homelessness, high bail rates, and the high price of mental health treatment.\textsuperscript{259,260,261,262}

\textsuperscript{252} U.S. Census Bureau, “Quick Facts,” U.S. Census Bureau, July 1, 2023. https://www.census.gov/quickfacts/fact/table/US.
\textsuperscript{253} United States Sentencing Commission, “2023 Demographic Differences in Federal Sentencing.”
\textsuperscript{254} Dholakia, “How the United States Punishes People for Being Poor.”
\textsuperscript{257} These are inflation adjusted figures from 2015.
\textsuperscript{258} Rabuy and Kopf, “Prisons of Poverty: Uncovering the Pre- Incarceration Incomes of the Imprisoned.”
\textsuperscript{259} Borrelli, “The Interplay of Mass Incarceration and Poverty.”
\textsuperscript{260} Hayes and Barnhorst, “Incarceration and Poverty in the United States.”
\textsuperscript{261} John Tierney, “Prison and the Poverty Trap.”
It is unlikely that these disparities are due to conscious, intentional bias of judges. However, risk assessment calculations are embedded in our country’s history, which, unfortunately, includes considerable bias. Many of the metrics used in predicting criminality serve as proxies for race and socioeconomic status—a relic of a criminal justice system with roots in slavery. Examples of such metrics are zip code, levels of neighborhood crime, money saved, number of parents and friends who have been incarcerated, level of educational attainment, grades, and more. Our country’s history has produced a state of affairs in which these metrics are predictive not only of criminality, but of race and socioeconomic status—even though the data points alone are not essentially related. If the data we have to train this model is biased in this same way, it seems impossible to not encode into it the very bias we hope to avoid. A related concern is that when we improve our systems with such efficient new models, it is unlikely that we will change them. The reason for this resistance to change is financial.

263 Davis, “Are Prisons Obsolete?”
265 Harcourt, “Risk as a Proxy for Race: The Dangers of Risk Assessment.”
273 Hao, “AI Is Sending People to Jail—and Getting it Wrong.”
274 Heilweil, “Why Algorithms Can Be Racist and Sexist.”
O’Neil argues that as long as our metric of success for AI models is cost-efficiency, we are unlikely to change them.\textsuperscript{276} I will expand upon this later.

Opacity

Are the algorithms that are used in courtrooms opaque? In two major ways, yes. First, the COMPAS program that was used in Eric Loomis’ case in Wisconsin was created by a company called Northpointe, Inc.\textsuperscript{277,278,279,280} While the questions that are asked on the survey are publicly available, the algorithm itself remains a trade secret.\textsuperscript{281,282,283} Therefore, in a very literal sense, the algorithm is opaque because judges, clerks, lawyers, defendants, and the public, are unaware of exactly how it makes its prediction.

The second way in which these algorithms are opaque is that they may become so complex that humans have difficulty understanding how they reach their predictions.\textsuperscript{284,285,286,287,288} Earlier I discussed how predictive algorithms are able to accommodate the exceptionally complex relationship between an individual’s personal data and criminality. In my example, I noted that the relationship between two variables, A and B, led us to make a

\begin{itemize}
  \item O’Neil, “Weapons of Math Destruction.”
  \item Now called equivant.
  \item Smith, “In Wisconsin, a Backlash against Using Data to Foretell Defendants’ Futures.”
  \item Angwin, “COMPAS Assessment Questionnaire.”
  \item Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”
  \item Ibid.
  \item Smith, “In Wisconsin, a Backlash against Using Data to Foretell Defendants’ Futures.”
  \item Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis.”
  \item Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”
  \item Rudin & Radin, “Why Are We Using Black Box Models in AI When We Don’t Need to? A Lesson from an Explainable AI Competition,” \textit{Harvard Data Science Review} 1, no. 2 (November 1, 2019). https://doi.org/10.1162/99608f92.5a8a3a3d.
  \item O’Neil, “Weapons of Math Destruction.”
\end{itemize}
certain prediction, and then added a third, C, to show how the introduction of new variables changes the relationship between the first two and our predicted outcome. This process can continue to get staggeringly more complex. We can take into consideration D, E, F, G, H, I, and so on, until we have a highly detailed picture of the individual’s entire life. To provide an example of how nuanced this can be, studies show that unaddressed vision problems are associated with greater rates of juvenile recidivism.\(^{289}\) As we add each new variable, a prediction about their future behavior will change, either slightly or significantly. AI-powered programs will be able to handle these changes without issue, cross-analyzing the change in relationship between the individual’s many data points and the data points of past offenders, as each variable is added. In theory, this is not difficult to follow.

However, research shows that humans are unable to make as good use of additional data inputs as statistical models—eventually humans begin to pick the data inputs that seem to them to have the most predictive power.\(^{290}\) This can hamper accuracy, especially in borderline or anomalous cases\(^ {291}\): if the inclusion of variable D, for example, changes the outcome significantly, but humans do not see this as a legally relevant factor, then they will neglect its impact.

AI’s ability to detect highly complex patterns in data creates problems for human interpretation in both directions. Did the algorithm’s prediction change because it has made a mistake? Or does it detect a meaningfully predictive relationship among the defendant’s many interconnected data points that humans are unable to see? Similar to the way many people would


\(^{290}\) Lin et al., “The Limits of Human Predictions of Recidivism.”

\(^{291}\) Ibid.
not feel confident questioning a program’s calculation about astrophysics, many people may feel uncomfortable questioning AI-powered algorithms’ predictions.292,293,294,295

This could result in algorithms being left unchecked.296,297 Its complexity gives it an air of scientific authority that can lead people to grant the model too much respect.298 This does not just happen to laypeople.299 Predictive algorithms can become so complex that even the data scientists who programmed the software do not understand why it gives the prediction it does.300 This will probably become truer as the technology becomes more powerful, potentially leading to false predictions in the sentencing process being passed off as scientifically justified, and thereby beyond reproach.301

Incomprehensibly powerful algorithms could restructure our criminal justice system in unforeseen ways. If a prison sentence is viewed as beyond reproach, this may also bring into question the role of the Criminal Appeals Court. Appeals Courts investigate whether the law has been fairly and accurately applied by the lower courts.302 But if we have technology producing such accurate predictions that humans are uncomfortable questioning them, what would be the

---

294 Bagchi, “Why We Need to See Inside AI’s Black Box.”
295 Blouin, “AI’s Mysterious ‘black Box’ Problem, Explained.”
297 Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”
298 Rodu & Radin, “Why Are We Using Black Box Models in AI When We Don’t Need to? A Lesson from an Explainable AI Competition.”
299 Blouin, “AI’s Mysterious ‘black Box’ Problem, Explained.”
300 Bailey, “Welcoming our new algorithmic overlords?”
301 Bagchi, “Why We Need to See Inside AI’s Black Box.”
point in appealing the decision? And if the appellate judge understands the prediction no better, then what are they to do about it anyway?  

Scale

Can this technology scale? This is a question of whether this technology, actually rather mundane when compared to some of the technology being released every day now, is capable of scaling into a “tsunami force, one that [will] define and delimit our lives.” I believe that I have shown that the answer to this is yes. Our criminal justice system is at the center of so many of the socioeconomic and political challenges our country faces, and impacts so many peoples’ lives, that it is simply hard to imagine a greater scale.  

Crossroads

This brings us to a crossroads for our criminal justice system. Our society is going to have to make some pretty severe and practical moral judgments in the coming years. Such powerful technology offers profound positive change. But there is also cause for great concern: if we make such a momentous commitment to improving the efficiency of our justice system, do we risk sealing in place aspects of it that are harmful to our society?

---

303 Rudin et al., “The Age of Secrecy and Unfairness in Recidivism Prediction.”
305 Hao, “AI Is Sending People to Jail—and Getting it Wrong.”
I argue yes. I endorse this technology’s creation and continued development, for I hope that it will remove from our sentencing process the element of quasi-scientific—almost spiritual—character assessment. However, I argue that the system in which this model will be placed is extremely harmful. We must make major, structural changes to our system so that the recidivism prediction model we implement does not function with the de facto biases in our system as guiding principles.  

In a Capitalist Economy, Cost-Effectiveness Rules

It is paramount that our criminal justice system takes these next steps deliberately—but also very quickly. This technology should undergo extensive testing and it is imperative that we ameliorate the features of our criminal justice system that cause significant damage to parts of our population before it is made available to judges. Failure to do so could be catastrophic. Thus, I argue that we must remove AI from our courtrooms until we have accomplished this reform and can be certain that these tools are more accurate and less biased than human judges.

However, the opposite development is currently underway—and I argue that the reason for this is not improved equality under the criminal justice system, but the cost-effectiveness offered by these models.  

In capitalist economies, cost-effectiveness is a key metric of

---

308 There is stern objection to the claim that our criminal justice system has the type of bias I have described here. Even among those who agree that our criminal justice system is problematically racist, for example, there is disagreement about the intentionality and source of that bias. I will address this later in this paper.

309 I want to make clear here that my argument is descriptive rather than normative. I don’t aim to make a value judgement here of capitalism’s regard for cost-effectiveness.
It refers to the ability to achieve the desired outcome at the lowest possible cost. This is a pattern that is readily observable across our economy: when employers are able to replace workers with cheaper labor without seriously sacrificing productivity, they tend to do so.

One might very reasonably object that our criminal justice system is not a business, it is a part of our government. However, many features of our government mimic the market economy. One of those features is the parallel between government officials promising voters and corporate executives promising shareholders that they will reduce costs through increased efficiency. So, while judges are government officials and quite unlike corporate employees, their employment remains vulnerable to very similar market forces.

---

313 Thompson, “Cost Effectiveness vs Cost Efficient: The Difference.”
321 Michaels, “Running Government like a Business . . . Then and Now.”
It has been suggested that AI may be the new “offshoring.”\textsuperscript{322,323} AI is already replacing humans in certain jobs, like budget analysis, web development, tax preparation, and technical writing.\textsuperscript{324,325} Although for many jobs AI remains more expensive than its human counterparts, it is only a matter of time before this changes.\textsuperscript{326,327,328} Simply put, I argue that our economy rewards the benefits of AI. As its speed and accuracy increase and its cost decreases, we will come to a point in which a simple risk-reward calculation will lead us to cede much of the control in sentencing deliberation to the machine.

Many people will immediately reject this prospect, arguing that technology is not infallible.\textsuperscript{329,330,331} Replacing humans in the role of risk assessment and sentencing is almost certain to lead to mistakes and systematic under- or over-sentencing. Moreover, studies show

\begin{itemize}
  \item \textsuperscript{327}Aaron Raj, “MIT Study: AI is Costlier than a Human Workforce,” Tech Wire Asia, January 25, 2024. https://techwireasia.com/01/2024/mit-study-humans-are-cheaper-employees-compared-to-ai-workforce-and-robots-%F0%9F%A4%96/#:~:text=Only%203%25%20of%20such%20tasks,still%20a%20more%20economical%20option
  \item \textsuperscript{328}A related topic is how this will happen for manual labor jobs. For many jobs, replacing the human worker is not as simple as creating a program capable of carrying out the mental processes of the worker. How would we replace construction workers or warehouse workers? This would be done with robots—which are not themselves artificial intelligence. Robots are simply a house for AI (Mark Talmage-Roston, “How Will Artificial Intelligence Affect Jobs 2024-2030”). There are efforts being made to create such robots (Kate Rogers, “The Robots Are Coming - and the Companies Building Them Are Looking for Workers.”), but this is an additional obstacle that must be overcome in order to replace workers. However, I am not arguing that judges will be replaced with robots, so this topic is outside the scope of my paper.
  \item \textsuperscript{330}Ludwig and Mullainathan, “Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System.”
  \item \textsuperscript{331}Rudin et al, “The Age of Secrecy and Unfairness in Recidivism Prediction.”
\end{itemize}
that AI can manifest the types of bias that we are hoping to avoid—racism, sexism, classism—called data bias.\textsuperscript{332,333,334} These are valid criticisms—in fact, they are the criticisms I have raised in order to show that risk assessment tools are likely to become WMDs—but they are also problems that already hamper our system.\textsuperscript{335,336,337,338} Moreover, these are moral arguments about the potential harms of this technology—but this is about money. The program doesn’t need to be perfect at recidivism prediction, case analysis, or free of bias. It just needs to be good enough that the benefits are greater than letting a human do the work—and the benefit of reduced cost to the state are heavily weighted by voters.

Consider the pressure of the taxpayer’s demand to cut costs.\textsuperscript{339,340,341} We need look no further than the creation of New Hampshire’s Circuit Court in 2011.\textsuperscript{342} In 2010, New Hampshire

\begin{footnotesize}
\begin{itemize}
  \item Harcourt, “Risk as a Proxy for Race: The Dangers of Risk Assessment.”
  \item Friis and Riley, “Eliminating Algorithmic Bias Is Just the Beginning of Equitable AI.”
  \item Heilweil, “Why Algorithms Can Be Racist and Sexist.”
  \item Davis, “Are Prisons Obsolete?”
  \item Reiman and Leighton, “The Rich Get Richer and the Poor Get Prison.”
  \item Georgia Innocence Project, “Beneath the Statistics: The Structural and Systemic Causes of Our Wrongful Conviction Problem.”
  \item National Registry of Exonerations, “Race and Wrongful Convictions in the United States 2022.”
  \item I use the state of New Hampshire because this is the state in which I have written my thesis. It should be noted that New Hampshire is a comparatively sparsely populated state, so the cost analysis I run here could be considerably greater for a more densely populated part of the country.
\end{itemize}
\end{footnotesize}
faced a budget crisis. Governor John Lynch, under pressure to reduce government spending, urged Chief Justice John Broderick to form an Innovation Commission. The commission suggested the Circuit Court, which combined the District, Family, and Probate Divisions and gave all Circuit Court judges the authority to handle all three types of cases. “The concept was simple: combine three limited jurisdiction courts, collapse management structures, and certify judges to hear all cases. House Bill 609 was introduced with bipartisan support in January 2011 and, with lightning speed, signed into law. The Circuit Court began operations July 1, 2011, and has since managed over 1.2 million new cases.” With this change, New Hampshire reduced its number of clerkships from 118 to 18 (and it is now 16): “In March 2011, 118 clerks and deputy clerks were told they could apply for any of 52 new positions that would be posted, including 18 new clerk positions; anyone who didn’t get a job would likely be laid off. In a few painful short weeks, we went from a structure of 52 clerks down to 18 who manage all Circuit Court divisions in 35 courthouses around the state. The immediate savings in management payroll was nearly $2.1 million the first year and, unadjusted for increased salaries

---

347 King, “The Circuit Court after 9 Years.”
349 King, “The Circuit Court after 9 Years.”
350 Ibid.
and benefits, is nearly $19 million over the first nine years.”\textsuperscript{352} Those first nine years also saw a reduction in Circuit Court Judges, from 40 full-time and 29 part-time judges to 2 marital masters, 34 full-time and 8 part-time judges.\textsuperscript{353}

The article in the New Hampshire Business Review that outlines this process proudly shows that this reform saved New Hampshire taxpayers $55 million over the next nine years.\textsuperscript{354} Governor John Lynch enjoyed bipartisan support and had one of the highest approval ratings for a governor in the country.\textsuperscript{355} Although this was mostly just a restructuring plan, the parallels should be apparent.\textsuperscript{356} The objective for the Innovation Commission was greater cost-effectiveness, which was produced by an increase in the responsibilities of Circuit Court judges and a decrease in the number of judges and clerks.\textsuperscript{357} I argue that a similar cost-benefit analysis is coming for judges as we become able to make judicial deliberation more efficient.

With judges able to get through cases more quickly, we will be able to reduce the backlog of cases that our country currently suffers from. As we are able to better manage America’s legal caseload, we will need less judges—an opportunity to reduce operational costs that we will certainly take. We can get a clearer view of this cost-benefit analysis by weighing the costs to our state’s government of paying judges and clerks versus using a tool like COMPAS. In the state of New Hampshire, there are three court systems after the merger in 2010: the Supreme, Superior, and Circuit Courts. There is one Supreme Court in the state, which handles appeals

\begin{flushright}
\textsuperscript{352} King, “The Circuit Court after 9 Years.”
\textsuperscript{353} Ibid.
\textsuperscript{354} Ibid.
\textsuperscript{356} The plan did also include an automated motion-filing system, which enabled attorney and self-represented litigants to upload their motions rather than filing in person. This allowed a small reduction in administrative personnel on the payroll (King, “The Circuit Court after 9 Years.”).
\textsuperscript{357} King, “The Circuit Court after 9 Years.”
\end{flushright}
from the Superior and Circuit Courts. The New Hampshire Supreme Court is composed of a Chief Justice and four Associate Justices. The Chief Justice earns $204,076 per year, while each Associate Justice earns $197,937. The Superior Court is composed of eleven Superior Courthouses throughout the ten New Hampshire counties, with a total of twenty Superior Court Justices. In the Superior Court, there is one Chief Justice, who earns a salary of $197,937, and nineteen Associate Justices, who earn salaries of $185,638. In the Circuit Court, there are two Administrative Judges, who each earn $197,937, and forty-four Associate Judges, who each earn $185,638. Between all of these judges, New Hampshire pays $13,284,829 per year in salaries. Additionally, there are sixteen judicial clerks total through the three court systems in New Hampshire. Each one is paid $61,562.80, unless it is their second year of clerking, for which they are paid $64,643.80. Assuming that each of the sixteen clerks is a first-year clerk, every year the state of New Hampshire pays approximately $985,000 in clerkship salaries, bringing the annual salary total to $14,269,829.

Now consider the cost of COMPAS: approximately $68,000 for the first year and $19,000 for maintenance and support for each subsequent year. Programs like COMPAS will also continue to get more accurate, faster, and will undoubtedly soon combine the abilities

361 New Hampshire Law Library, “Title LI Courts - Chapter 491A: Judicial Salaries.”
362 New Hampshire Law Library, “Title LI Courts - Chapter 491A: Judicial Salaries.”
364 New Hampshire Law Library, “Title LI Courts - Chapter 491A: Judicial Salaries.”
366 This is an estimate because Northpointe Inc. (now “equivant”) does not make the price of COMPAS available to the public. I was unable to find any resources on this other than the San Benito Country Agenda Item.
of generative AI like ChatGPT. With this available to a judge, the deliberative process can be reduced to hours.\textsuperscript{367,368} With the ability to have a decision drafted within minutes, the need for judicial clerks disappears. Not only would this be cheaper for the court system, but it would also likely make things much more efficient. With greater efficiency in the court system, the greater our ability to reduce the number of judges on the payroll.\textsuperscript{369}

I argue that, ultimately, the impetus for such tremendous change will not be an attempt at improved judicial outcomes—it will be regular old budget cuts. Eventually, the governor is going to come knocking on the chief justice’s door again.

We Must Remove This Technology from Our Courtrooms Now

Perhaps our criminal justice system is less at a crossroads than we are driving dangerously quickly down the wrong road. This technology has the potential to cause major harm to our society if we fail to first make substantial changes to our criminal justice system—and it has already been implemented in many courtrooms across the country.\textsuperscript{370} I do not believe that we can make the changes we need to as we increase our reliance upon AI. This is because of its symbiotic relationship to the system—it will learn from our harmful system and thereby recreate it with incredible efficiency. Additionally, as we remove the human employees that previously did this work, we will reduce our ability to turn back without majorly compromising the efficiency we use as a metric of success. I believe that such a reduction in efficiency twenty

\textsuperscript{367} Ortiz, “How ChatGPT (and Other AI Chatbots) Can Help You Write an Essay.”  
\textsuperscript{368} Monterey Herald, “ChatGPT, the Good AI, EssayGenius: Can They Write My Essay?”  
\textsuperscript{369} King, “The Circuit Court after 9 Years.”  
\textsuperscript{370} Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis.”
years from now will make AI segregation a non-starter—even if our system creates much greater harms.

Seismic economic forces act on tech innovation and integration. Moreover, they prevent us from winding back the clocks to a time of relative inefficiency. Thus, I argue that we must remove this technology from our courtrooms across the country right now. If we fail to take action now, we effectively make a decision in the direction of greater integration—and therefore greater harm.

Objections

There are many reasonable objections to the arguments that I have made. I will address some of them here. First, one might object that my argument commits the slippery slope fallacy. A slippery slope argument is one in which “a course of action is rejected because, with little or no evidence, one insists that it will lead to a chain reaction resulting in an undesirable end … The slippery slope involves an acceptance of a succession of events without direct evidence that this course of events will happen.” I have argued that because judges have begun to use AI to inform or determine sentencing, we are on a path towards AI taking the sentencing role entirely from human judges. Because of the enormous harm this may cause, I argue that we should remove this technology from courtrooms.

---

I do not believe that my argument constitutes a slippery slope fallacy. A degree of speculation will be present in almost all decisions about how to use AI. The technology is powerfully intelligent in a way that humanity has not yet encountered—and will rapidly continue to get more powerful as we work our way towards creating AGI. In order to meaningfully discuss whether it is a good idea to begin using this rapidly developing technology in a given space, we have to make some educated guesses about what its future might look like.

It is true that I do not know that AI will take the sentencing role from human judges, but I believe that I have provided sufficient evidence to reasonably assert that it may. I argue that the likelihood of its happening is enough to take action to prevent its happening. This is because of the oppressive harm this could cost our society. Ironically, this is the same reasoning pattern used to justify the process under review: we justify risk assessment because if there is a high level of risk that an offender is likely to harm others in the future, we want to take action about that now. Although the prediction technology—or human judge—will never know that the offender will commit another crime, we accept that a certain level of risk is sufficient evidence to act.

---

376 Bostrom, “How Long Before Superintelligence?”
379 Skeem and Lowenkamp, “Risk, Race, and Recidivism: Predictive Bias and Disparate Impact.”
381 Neufeld, “In Defense of Risk-Assessment Tools.”
382 Ibid.
383 Monahan and Skeem, “Risk Assessment in Criminal Sentencing.”
384 Berk and Bleich, “Statistical Procedures for Forecasting Criminal Behavior.”
A second objection to my project might be that AI does not replace human workers in the way that people fear.\(^{385,386}\) Instead, it enables humans to do the better part of their work.\(^{387}\) One way this argument has been supported is by highlighting the way that Warren Buffett has used natural intelligence (human employees) in this way for decades.\(^{388}\) Buffett employs many people to do the menial tasks of running the numbers and writing reports on company valuation and forecasting.\(^{389}\) This is, fundamentally, non-technological workplace automation.\(^{390}\) Because he has outsourced all this work to employees, he is able to spend all his work time reading and prognosticating in search of his next profitable stock pick.\(^{391}\) This objection suggests that AI will enable most jobs to be enhanced in this way, automating the menial parts of the job and allowing humans to do the higher-order elements.

There is indeed a structural similarity between what Warren Buffett has done with human employees and how artificial intelligence can be used to make a workplace more efficient. However, I think that this objection has missed something quite obvious: the jobs of those human employees who will have their roles eliminated by technological automation. Buffett can now replace many of those human employees with artificial intelligence—and if he aims to increase productivity and decrease spending, he should. So, for Warren Buffett and anyone else who owns and runs an incredibly successful business, this is a correct analysis of the way that AI can

\(^{388}\) Ibid.
\(^{389}\) Ibid.
\(^{391}\) Weisenthal et al., “The Economist Who Believes AI Will Be Great for the Middle Class.”
help to enhance their work rather than replace them. However, the employees that have been replaced have not had their jobs enhanced. They have had them eliminated.

If Buffett fired all of his lower-level employees and replaced them with AI, those employees would obviously have to find other jobs. However, if all investment companies like Berkshire-Hathaway make a similar move, then those analysts must find something else to do entirely. A capitalist economy requires that one works in order to live—shelter and food costs money.\(^\text{392}\) For a time, pivoting careers will be possible, but as our reliance upon AI increases, there will be less companies willing to spend the extra money on less-efficient human employees.\(^\text{393,394}\)

There has been considerable discussion about how these events might unfold.\(^\text{395,396,397}\) Many fear that this will produce such a wealth gap between a small group of ultra-wealthy titans of industry and everyone else that society will collapse.\(^\text{398}\) My goal is not to argue that this potentially dire shortage of jobs will occur in any particular way, but to suggest that this objection—that AI will not replace many jobs in the future—is naive. For judges specifically, I argue that AI will enhance the work of those who remain on the bench—but that it will also reduce the number of judges and clerks significantly. So, this objection makes sense for some


judges, but certainly not all. Moreover, AI doing the risk assessment and sentencing seems to be taking the “higher order” functions of the job from the judge, leaving the judge to do administrative tasks like case management.

Another objection to my paper regards the presumption that racial and socioeconomic factors should not be considered in sentencing. **What if race and socioeconomic status are actually the best predictors for recidivism? Should we not make that information available to prediction models just so that we do not offend liberal worldviews?**

Arguing that we should not use certain data points because we think it is racist or sexist seems to insert a subjective opinion into the equation. Is that not what we are hoping to avoid with using predictive algorithms in the first place?

These are questions that many people may wonder, but few want to ask aloud. They are therefore important questions. In the United States, it is actually true that race and socioeconomic status are statistically correlated with criminality. However, this is not because certain shades of human skin produce criminal characteristics in an individual, or because people who live in poverty inherently enjoy committing crime for its thrill. This is instead because (1) a life of poverty produces certain circumstances that contribute to criminality, namely financial

---

399 Probably those which are at the top of the socioeconomic hierarchy.
404 Skeem and Lowenkamp, “Risk, Race, and Recidivism: Predictive Bias and Disparate Impact.”
desperation and untreated mental illness, and (2) there exists a correlation between race and poverty due to this country’s history of slavery. 406,407,408,409

In the United States, people of lower socioeconomic classes—and therefore often those with Black and Brown skin410,411—are systematically disadvantaged to the extent that many experience despair: “a sense of being trapped in a world where one’s sense of self is undermined, dignity is compromised, and few options exist for a way forward.”412 Additionally, the high cost of healthcare means that mental health crises are often left untreated.413 These are both phenomena that have been shown to lead people to commit crimes.414,415,416 The criminal justice system then uses these connections to systematically impose harsher punishments, thereby reifying the correlation.417 Because the country’s history has both produced this connection and thenpunishes for it, many see these criteria’s inclusion as a bad faith maneuver.418,419,420

---

407 Wiggins, “Proxies.”
410 Hipps, “The Effect of the Distribution of Income and Race/Ethnicity Across Households and Neighborhoods on City Crime Trajectories.”
414 Hipps, “The Effect of the Distribution of Income and Race/Ethnicity Across Households and Neighborhoods on City Crime Trajectories.”
417 Smith, “The Cyclical Nature of Criminal Offending: A Qualitative Exploration of Intermittency among Active Criminal Offenders.”
418 Heilweil, “Why Algorithms Can Be Racist and Sexist.”
419 Hao, “AI Is Sending People to Jail—and Getting it Wrong.”
420 Matos and Hodge, “The Chains of Slavery Still Exist in Mass Incarceration.”
This is why many people are concerned about this technology: it does not “know better.” It learns from the data that is fed to it, and it knows no other way. Rather than making a nuanced assessment of an individual’s disadvantaged upbringing and deciding on this basis that he deserves a break, it identifies this tough upbringing as itself a reason to impose a harsher sentence. This is so shallowly statistics-oriented that it robs the process of compassion and contradicts the value of fairness that we wish for our criminal justice system. Furthermore, its complexity passes off internalized systemic bias as scientific objectivity, preventing progress rather than helping to achieve it.

At this time, it is impossible to use either of these factors without invoking our country’s history of slavery and racism. While we might informally think of risk assessment as something like a calculation of the “badness of the individual,” it is a complex calculation including all the contextual forces that have acted upon the given individual for his entire life. Our hope is that artificial intelligence will help us get closer to making this a more scientifically sound endeavor, but training it with faulty correlations like the one that exists between race, poverty, and criminality, will produce a feedback loop that perpetuates this bias in our system.

This leads to another objection concerning my seeming absolution of offenders’ responsibility for their crimes. I appear to be saying that if someone has been sufficiently acted upon by contextual forces, they are not responsible for their crime. I made a similar claim earlier in my paper by arguing that our criminal justice system commits unjust harms by pulling

424 Bushway, “‘Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
impoverished people into it. **I describe a situation in which criminals are somehow not responsible for their own behavior. This seems impossible—humans have free will.**

I push back against the charge that I have absolved people of responsibility for their actions. There are, however, degrees of culpability, which we already account for in our criminal justice system. Rather than claiming that people from impoverished neighborhoods are completely out of control of their own actions, I argue that there are socioeconomic forces that create circumstances of desperation and reduced ability to address mental health issues. Poverty can lead to low levels of hope, high levels of despair, and mental health crises, which in turn are major contributors to criminality.

Our criminal justice system is so intricately connected to the other systems in our society that any harmful features of the criminal justice system reciprocally produce harms in other systems. For example, when a father goes to prison, his children are impacted by the absence of their father—both psychologically and economically. This can make life harder in

---

425 The claim that humans have free will is such a complex, hotly contested debate that I simply do not have space to address it in this paper.
428 Knifton and Inglis, “Poverty and Mental Health: Policy, Practice and Research Implications.”
429 Offenheiser, “Inequality, Dignity, and Despair.”
432 Psychologically through the distress of having an incarcerated parent, and economically through the loss of one parent’s income.
434 Wildeman, “Parental Imprisonment, the Prison Boom, and the Concentration of Childhood Disadvantage.”
a way that can impact the children’s performance in school. Poor performance in school and reduced parental guidance make the children less likely to complete high school, which reduces employment prospects. When someone is unemployed, they have no money with which to pay for the necessities of life. This produces a desperate situation in which people sometimes turn to criminal activity to get by. Evidence suggests that many who commit crimes do so out of necessity—debt, hunger, threat of violence, et cetera. Criminal activity, obviously, leads to a greater likelihood of becoming involved in the criminal justice system. Once involved in the criminal justice system, it is statistically more likely that one will remain involved—and this is what is meant by the cyclical nature of America’s prison system.

Even without some of the more pressing circumstances of dire poverty, a low income can push people towards the prison system. Consider the increased prevalence of plea bargains. The U.S. Department of Justice states that “A defendant may only plead guilty if they actually committed the crime and admits to doing so in open court before the judge. When the defendant admits to the crime, they agree they are guilty and they agree that they may be ‘sentenced’ by the

---

436 Ibid.
439 Unnever et al., “Public Support for Attacking the ‘Root Causes’ of Crime: The Impact of Egalitarian and Racial Beliefs.”
442 Smith, “The Cyclical Nature of Criminal Offending: A Qualitative Exploration of Intermittency among Active Criminal Offenders.”
Plea bargains are ideally viewed as a way to save time and money by allowing the guilty to admit to their crime and save everyone involved the trouble of spending months or years in a trial. However, a recent report from the American Bar Association states that innocent people are often coerced into plea bargains by pressure from prosecutors, defense attorneys, and the daunting prospect of decades-long mandatory minimum sentences. And indeed, trial sentences are on average three times longer than plea bargain sentences.

A legal system that incentivizes plea bargains encourages disparate outcomes for the privileged and disadvantaged. Those who are financially able hire private criminal defense attorneys who will fight their case in court, whereas those who are unable to afford a private attorney receive a public defender. Public defenders offices are notoriously understaffed and

452 Ibid.
453 Hessick et al., “Plea Bargains: Efficient or Unjust?”
overloaded with cases. These limited resources mean that public defenders—especially compared to the considerably more funded state’s attorneys—are more likely to encourage clients who have a difficult case to take a plea deal in exchange for a reduced sentence. This is cause for serious concern. In addition to sacrificing the democratic process of a jury trial in the name of cost-efficiency, over 20% of exonerations are for convictions based on guilty pleas. A defendant’s false guilty plea has a direct impact on recidivism prediction and sentencing in the future. If the defendant is lying about having committed a crime in order to receive a light prison sentence, this produces bad data. All the defendant’s data has now officially been statistically correlated with the commission of a crime that the defendant did not commit. Since conviction history is part of the assessment, this incorrect data is then used for future recidivism prediction for that individual if they are ever

457 Fair Trials Organization, “Plea Bargaining.”
458 It should just be explicitly noted that public defenders do not necessarily do this because they are simply overloaded with cases and can’t be bothered. Sometimes, perhaps even oftentimes, a public defender might believe that their client is innocent but also know that they are very unlikely to win in court.
462 Smith, “The Cyclical Nature of Criminal Offending: A Qualitative Exploration of intermittency among Active Criminal Offenders.”
charged with a crime again—which evidence shows is likely.\textsuperscript{463,464,465,466} This false correlation is then also used for other peoples’ recidivism assessments.\textsuperscript{467,468} And since familial criminal history is part of the risk assessment, the false conviction is now a significant factor in the defendant’s children’s assessment if they are ever charged with a crime.\textsuperscript{469,470,471,472} This is how innocent people get caught in the “revolving door” of the criminal justice system.

So, yes, when an individual has committed a crime, on some level they were mentally present and made a choice to commit the crime\textsuperscript{473}—but many people have also been led to this decision through a life of systematic disadvantage. I argue that these people should be considered less guilty than those who have committed a crime out of interest or pure greed—for example, wealthy people who commit insider trading.\textsuperscript{474}

\textbf{Another objection might be to my characterization of our criminal justice system, which sounds somewhat malicious.} The criminal justice system is meant to produce justice. Why have I portrayed it as such a harmful system?

I do not argue that the system is malicious—simply unjustly harmful. One does not need to intend to cause harm in order to do so. Jeffrey Reiman and Paul Leighton, in their book \textit{The

\begin{itemize}
\item \textsuperscript{463} Bushway, “‘Nothing Is More Opaque than Absolute Transparency’: The Use of Prior History to Guide Sentencing.”
\item \textsuperscript{464} Harcourt, “Risk as a Proxy for Race: The Dangers of Risk Assessment.”
\item \textsuperscript{467} Skeem and Lowenkamp, “Risk, Race, and Recidivism: Predictive Bias and Disparate Impact.”
\item \textsuperscript{468} Davies and Douglas, “Learning to Discriminate: The Perfect Proxy Problem in Artificially Intelligent Criminal Sentencing.”
\item \textsuperscript{469} Yong, “A Popular Algorithm Is No Better at Predicting Crimes than Random People.”
\item \textsuperscript{470} Wildeman, “Parental Imprisonment, the Prison Boom, and the Concentration of Childhood Disadvantage.”
\item \textsuperscript{471} Bryant, “Children Suffer When Parents Are Imprisoned.”
\item \textsuperscript{472} Martin, “Hidden Consequences: The Impact of Incarceration on Dependent Children.”
\item \textsuperscript{473} Unless they have been found to have reduced culpability (Jan Willem Wieland, “Degrees of Criminal Culpability”).
\item \textsuperscript{474} However, it is sometimes framed the other way, as in cases like Ethan Couch’s, who was found to be less culpable of his crime because he was so affluent that he could not appreciate that there are repercussions for his actions (Anthony Zurcher, “‘Affluenza Defence’: Rich, Privileged and Unaccountable.”).
\end{itemize}
Rich Get Richer and the Poor Get Prison, argue that the United States justice system’s de facto purpose is to serve the wealthy and protect them from the poor.\textsuperscript{475,476,477} The authors invite the reader to entertain the idea that “the goal of our criminal justice system is not to eliminate crime or to achieve justice but to project to the American public a credible image of the threat of crime as a threat from the poor.”\textsuperscript{478} Their aim is not to argue that the system was intentionally built this way, but that when we examine the system through this lens, many of its features make more sense.

Their argument goes as such: (1) our justice system fails to protect people from the crimes they fear by alleviating the poverty that causes them; (2) it focuses attention on the crimes of the poor (“street” crimes) while failing to protect people from the harms caused by the crimes of the wealthy (“white-collar” crimes)—and it also fails as much as possible to enforce the law against wealthy people who commit “street” crimes; and (3) it therefore succeeds in creating the image that crime is almost exclusively the work of the poor.\textsuperscript{479} This serves the interest of the powerful by mechanically weeding out the rich at every possible juncture throughout the system while also sweeping in as many poor people as possible by over-policing to alleviate the fear that the poor are causing so much crime.\textsuperscript{480} Regardless of intentionality, our criminal justice system has long caused systematic damage to minoritized racial groups in our country. These are unjust

\textsuperscript{475} Reiman and Leighton’s work on the nature of the American criminal justice system presumes a certain level of innocence—surely, they imply, no one would intentionally create such a harmful system. Thus, as a reasonable country, we should take the corrective measures they suggest in their book: decriminalizing illicit drugs and victimless crimes, actively prosecuting white-collar crime, reducing poverty, and giving all accused people access to high-quality, well-resourced defense attorneys.
\textsuperscript{476} Other writers, including Angela Davis, diverge from Reiman and Leighton’s perspective in arguing that the systemic harm caused by our criminal justice system is intentional and a direct result of its roots in American slavery.
\textsuperscript{477} Reiman & Leighton, “The Rich Get Richer and the Poor Get Prison.”
\textsuperscript{478} Ibid.
\textsuperscript{479} Ibid.
\textsuperscript{480} Ibid.
harms—regardless of the system’s intention—that stand to be intensified by the introduction of AI.

**One final objection to my paper might be that it is too late to remove AI from our courtrooms.** I argue that we are unlikely to make our systems less efficient because of our socioeconomic prepossession with cost-effectiveness. Since we are already improving the efficiency of our criminal justice system with AI models, doesn’t my argument mean that it is unlikely that we will remove this technology from our judges’ chambers because it will make things less efficient?

There is a difference between the two assertions—that it is unlikely, and it is too late. While I do worry that, for the reasons I have outlined in this paper, we are less likely to remove this technology from our courtrooms, I also believe that it is not entirely too late. I believe this, basically, because the general public seems to be ignorant to this development rather than supportive of it.\(^{481,482,483}\) The fact that many people seem to not know that judges are using AI programs in their sentencing deliberation suggests that this is a practice that has come about not as the result of widespread support but rather of largely unnoticed technological progress. I think this would be a lost cause if there were overwhelming public support for judges’ use of AI in their sentencing deliberation.

My hope is that when people learn about this development, they will begin to voice concerns to those who facilitate its implementation (and therefore could facilitate its removal):


\(^{482}\) O’Neil, “Weapons of Math Destruction.”

politicians, lawyers, and judges.\textsuperscript{484} If people begin to tell those in power that this is a political topic that matters to them and that they do not wish for this practice to be in place, it will encourage our political leaders to take action. This is how representative democracy works.\textsuperscript{485} This alone is obviously not enough, as philosopher Susan Anderson writes that what is required is “practical action to dislodge shared expectations, unsettle attitudes, and trigger practical deliberation. This requires concerted mass public action, effectively demonstrating a collective rejection of the entrenched norms—a dissolution of the prior shared expectations—and a determination to replace them with rival norms.”\textsuperscript{486} Nonetheless, the difficulty in accomplishing the removal of this technology from our judges’ chambers does not mean that it is too late to do so.

Conclusion

In this paper, I have explored some of the ways in which artificial intelligence might enhance the sentencing process through recidivism prediction technology. Notably, this technology can increase the accuracy of risk predictions and the speed with which sentencing decisions are reached. I then showed, however, that the recidivism prediction technology is likely to turn into what data scientist Cathy O’Neil calls a Weapon of Math Destruction. The potential harmfulness of this technology is due not to the inherent nature of the technology, but to the symbiotic relationship it will have with our already harmful criminal justice system.

\textsuperscript{484} Anderson, Elizabeth S.. “Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain’s Abolition of Slavery.” (2014).
\textsuperscript{486} Ibid.
O’Neil writes that, for AI models, “the heart of the problem is almost always the objective. Change that objective from leeching off people to helping them, and a WMD is disarmed—and can even become a force for good … To eliminate WMDs, … we must reevaluate our metric of success. Today, the success of a model is often measure in terms of profit, efficiency, or default rates.”

I have argued that the objective of implementing this technology is increased cost-effectiveness. It is against this metric that we will evaluate the technology’s success. Thus, if the technology makes our criminal justice system far more cost-effective—even if it proves to greatly increase harms done to society by the criminal justice system—we would be unlikely to substantially change our system once we have implemented it. Because of this, I argue that we ought to remove AI from our courtrooms now.

My project leads one to question: What sort of changes to the criminal justice system would be sufficient to allow the introduction of this technology to be safe? This is a question that I do not address here in great detail, but I hope to commit the rest of my career to properly answering it. I see this paper as a starting point for a much larger project: reforming our criminal justice system.

There are many changes that I believe would improve our system. Some ideas include reducing the over-criminalization of drugs, prosecuting white-collar crimes and street crimes with the same tenacity, reforming our education system, making it easier to find

---

employment once released from prison,\textsuperscript{491} taking a generally more rehabilitative approach to punishment,\textsuperscript{492} restructuring our policing system,\textsuperscript{493} bail reform,\textsuperscript{494} and reducing the prevalence of plea bargains so that we may alleviate some of the socioeconomic disparities caused by the different levels of legal representation afforded to wealthy versus poor defendants.\textsuperscript{495} Each of these are topics that could be (and are) written about at great length\textsuperscript{496,497,498,499}—but I suspect that sufficient change to our criminal justice system would include re-envisioning our country’s fundamental conception of what it means to achieve justice. These are changes that would require a cultural shift in addition to changes in the law.

\textsuperscript{496} Davis, “Are Prisons Obsolete?”
Bibliography

ACLU. “91 Percent of Americans Support Criminal Justice Reform, ACLU Polling Finds.”


https://doi.org/10.1371/journal.pone.0249454.


https://www.law.cornell.edu/wex/sentencing.


Enns, Peter K., Youngmin Yi, Megan Comfort, Alyssa W. Goldman, Hedwig Lee, Christopher Muller, Sara Wakefield, Emily A. Wang, and Christopher Wildeman. “What Percentage of Americans Have Ever Had a Family Member Incarcerated?: Evidence from The Family...


Algorithmic bias occurs when algorithms, criminal justice, and credit scoring.


https://www.brennancenter.org/our-work/analysis-opinion/mass-incarceration-has-been-driving-force-economic-inequality.

https://scholarship.law.cornell.edu/facpub/917


Johnson, Kevin. “‘Invisible’ Crisis: Public Defenders Still Underfunded, Understaffed 60 Years after Key Scotus Ruling.” USA Today, March 17, 2023.


https://doi.org/10.1037/lhb0000498


https://doi.org/10.1192/bjb.2020.78.

Knowledge at Wharton. “Your Data Is Shared and Sold... What’s Being Done about It?” Knowledge at Wharton, October 28, 2019.


https://nij.ojp.gov/topics/corrections/recidivism.

University of Michigan Law School, September 23, 2022.


Rudin, Cynthia, and Joanna Radin. “Why Are We Using Black Box Models in AI When We Don’t Need to? A Lesson from an Explainable AI Competition.” *Harvard Data Science Review* 1, no. 2 (November 1, 2019). https://doi.org/10.1162/99608f92.5a8a3a3d.


https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831279/.

Williams, Andrew. “Top AI Software Platforms and Their Use Cases.” Forbes, April 1, 2024.


Wisconsin Department of Corrections. “Compas.” DOC COMPAS, n.d.
https://doc.wi.gov/Pages/AboutDOC/COMPAS.aspx.


Xu, Zichun. “Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities.”
Applied Artificial Intelligence 36, no. 1 (December 8, 2021).

https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1107&context=public_law_and_legal_theory.
