Generating Bias: Why Artificial Intelligence is Not an Intelligent Choice for Jury Selection

Sarah A. Ferguson
3L, University of Pittsburgh School of Law

I.   Introduction

After a three-day jury trial, criminal defendant Miguel Angel Peña-Rodriguez was found guilty of unlawful sexual contact and harassment.[1] Following the trial, two jurors stayed behind, asking to speak with the defense counsel privately.[2] During this conversation, the jurors revealed that throughout jury deliberations, one juror expressed many anti-Hispanic views to influence the other jurors.[3] These views included statements about how he "believed the defendant was guilty because, in [his] experience as an ex-law enforcement officer, Mexican men had a bravado that caused them to believe they could do whatever they wanted with women," that "nine times out of ten Mexican men were guilty of being aggressive toward women and young girls," and that the defendant's witness was not credible because he was "an illegal."[4]

While the Supreme Court ultimately reversed and remanded this case in 2017, following an investigation into these comments, a question arises from these results – could this result have been avoided with advanced screening during jury selection? And with the growing abilities of artificial intelligence (AI) software, is AI the advanced screening solution needed?

Many believe that AI might be the future of jury selection, at the very least to be used to help attorneys gather more information about prospective jurors. In Miguel's case, if the defense attorneys had access to AI tools, they might have been able to find social media posts or online affiliations from the discriminatory juror prior to the trial and used that information to strike him.

---

[1] Peña -Rodriguez v. Colorado, 580 U.S. 206, 212 (2017).
[2] *Id.*
[3] *Id.*
[4] *Id.* at 212-13.

Even without blatant expressions of the juror's opinions online, AI might have been able to analyze data and trends about the juror, predicting any biases he might have had.

But even with these potential benefits, the risk AI poses in its current state is just too high to consider using in jury selection. While it may be more efficient, AI also poses an independent risk of bias, has a known issue with generating false information, and carries a high rate of public distrust due to privacy concerns. The following essay will analyze each of these risks independently, and show that AI, at least as it currently exists, cannot be trusted with a matter as fundamentally important to the fabric of our judicial system as jury selection.

II.   Advantages of AI Use in Jury Selection

It would be a disservice to the advancements of artificial intelligence to discuss the risks surrounding its use without first mentioning the potential advantages. Despite the controversy surrounding AI usage, it is undeniable that there are many ways AI could aid in increasing efficiency and fairness in jury selection.

The primary advantage of AI usage in jury selection is that AI can learn more information about a potential juror in a faster time than any attorney. Using a potential juror's online presence to inform jury selection is nothing new. Over the past two decades, as social media usage has become increasingly popular, attorneys have turned to online research to learn as much as possible about potential jurors before selection.[5] This process has rapidly changed, however, as the internet grows and improves. In 2012, one article called using the Internet to search for juror information a "needle in a haystack-type exercise," given how time-consuming and often unfruitful using the Internet to search for individuals was.[6] Now, thirteen years later, AI tools can

---

[5] Sonia Chopra, *Using the Internet and Social Media in Jury Selection*, PLAINTIFF MAG., Feb. 2012 https://plaintiffmagazine.com/images/issues/2012/02-february/reprints/Chopra_Using-the-Internet-and-social-media-in-jury-selection_Plaintiff-magazine.pdf.
[6] *Id.*

pull information about an individual from hundreds of sources in seconds. By using these tools, attorneys could access demographic information, social media activity, and even previous jury history instantly and use algorithms to assess this data for patterns, identifying potential biases and risks faster than ever before.[7] Recently, the legal AI company, Voltaire, launched a new tool specifically for jury selection, which searches for data related to a prospective juror, correlates the data using known patterns of human behavior, and generates a profile with the individual's expected personality type and viewpoints.[8]

In a case like Miguel's this generated profile could have pin-pointed the juror's racial bias, leading to informed jury selection and resulting in a fair trial the first time around. The potential power that this software holds in the future of Sixth Amendment fair trial rights is incredible and extremely tempting. However, as will be discussed throughout this paper, the risks that AI tools still pose are just too high.

III. Disadvantages of AI Use in Jury Selection

a. Bias within AI Models

While AI tools seem to have a future in identifying bias in potential jurors, one of the biggest disadvantages of using AI in jury selection is the risk of bias that exists within the model itself. It is easy to view AI insights as an all-knowing, neutral source, but as AI continues to be released to the general public, more and more bias in its analysis is revealed. AI databases and chatbots are not content-neutral but rather reflect the conditions and conceptions of the world in which they were created.[9] Intelligence systems are trained using real-world data sets but often ignore the

---

[7] *AI and the Future of Jury Trials,* CLAIMS & LITIG. MGMT ALL., Oct. 18, 2023, https://www.theclm.org/Magazine/articles/ai-and-the-future-of-jury-trials/2731.
[8] Mike Robinson, *How AI is Helping with Jury Selection and Why Some People are Concerned,* INFOTRACK, Feb. 13, 2023, https://www.infotrack.com/blog/ai-jury-selection/.
[9] Felix Dengg, *Biases in AI: How Neutral is Technology?,* DIGIT. GLOB., June 14, 2023. https://www.bmz-digital.global/en/biases-und-ki-wie-neutral-ist-technologie/.

nuanced issues and biases that already exist within this data.[10] In many ways, AI tools are simply reinforcing the bias that already exists, and packaging the information as neutral fact.

Examples of this bias can be seen across several different AI chatbots and databases, even within their short release. Several tests of AI models have revealed a racial bias within the systems. In a 2024 study on AI language models, researchers found that the models had a prejudice against certain dialects, such as African American Vernacular English, which leads to potentially harmful consequences about the predictions and patterns AI models will draw for these individuals.[11] In one hypothetical, when asked to pass judgment on a defendant who had committed first-degree murder, the AI model was significantly more likely to recommend the death penalty against defendants who spoke African American Vernacular English as opposed to Standardized American English, even without being overtly told the race of the defendant.[12] This bias against dialects poses a particular risk for use in jury selections when the attorneys would be looking at factors such as social media activity, allowing AI tools to profile individuals based on their speech and present these unwittingly bias judgments without explanation. The advantage of this instant data processing quickly becomes one of the biggest disadvantages when the generated result furthers harmful stereotypes. A similar study conducted in 2023 saw the damaging effects of these stereotypes when it examined language models integrated into healthcare systems and found that all of the models tested had examples of perpetuating harmful and inaccurate race-based medicine.[13]

---

[10] *Id.*

[11] Valentin Hofmann et. al., *Dialect Prejudice Predicts AI Decisions About People's Character, Employability, and Criminality,* CORNELL U., Mar. 2024, https://arxiv.org/abs/2403.00742.

[12] *Id.* at 3.

[13] Jesutofunmi A. Omiye, *Large Language Models Propagate Race-Based Medicine,* NPJ DIGIT. MED., Oct. 20, 2023, https://www.nature.com/articles/s41746-023-00939-z.

Racial bias is already a large concern in the jury selection process. Even after the

Supreme Court ruling in *Batson v. Kentucky* in 1986, which determined that the use of

peremptory challenges to exclude potential jurors based on race was a violation of the Equal

Protection Clause,[14] modern courts are still struggling to battle racial discrimination in jury

selection.[15] To use AI tools in their current state, knowing the struggles developers are still

having with eliminating this racial bias, would be irresponsible and inequitable.

The known biases in AI models go beyond racial bias as well. A 2024 study into

generative AI found that, when asked to produce narratives about individuals based on prompts,

AI models were four times more likely to describe women in domestic roles compared to their

male counterparts, with the only difference in prompting being gender.[16] The same study found a

similar truth when the AI model was told that an individual was gay, resulting in the AI models

producing negative narratives more often than positive ones.[17] Based on these studies, there is a

high likelihood that serious bias exists in AI models, for both these demographic traits and

potentially many more that have not yet been tested. Due to this risk factor, the danger of using

AI in jury selection is simply too high and would result in unnecessary pruning of diverse groups

from the jury pool.

b. *Generation of False Information*

In addition to AI models following bias, there is also a significant risk of AI models

generating false information, known as "hallucinations". While AI might be able to gather

---

[14] Batson v. Kentucky, 476 U.S. 79, 98 (1986).
[15] James E. Coleman Jr., *The Persistence of Discrimination in Jury Selection: Lessons from North Carolina and Beyond,* NAT'L ASS'N CRIM. DEF. LAWS., June 2018, at 28, https://www.nacdl.org/Article/June2018-ThePersistenceofDiscrimination.
[16] *Generative AI: UNESCO Study Reveals Alarming Evidence of Regressive Gender Stereotypes,* UNESCO, Mar. 2024, https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes.
[17] *Id.*

information about an individual in a larger quantity and faster than an attorney, it is highly

possible that the AI model would include generated, false information.

As AI has grown in popularity, countless examples of this phenomenon have made their way

to the public eye. Even in the legal world, examples of these hallucinations have begun to

surface. In 2023, attorneys in New York faced sanctions after using ChatGPT to conduct legal

research when the tool produced generated, fake quotes from non-existent legal cases.[18] More

recently, President Donald Trump's former attorney, Michael Cohen, faced public backlash after

he included fake case citations in an official court filing generated by the AI tool Google Bard.[19]

The generation of false information is a big enough problem in itself, but these examples

from legal cases highlight another issue with the use of AI tools – who is held accountable for the

failings of AI? In the two cases described above, the lawyers faced sanctions for their use of fake

case law, but would attorneys assume a similar burden for generated information about potential

jurors? If so, attorneys would be expected to fact-check each piece of information collected

about a potential juror, which undermines any perceived benefit of collecting data faster or more

efficiently.

Further, if an attorney were to go so far as to ask an AI tool to actively make decisions about

who the best jurors would be, rather than just providing information about the jurors for the

attorney to decipher, the "black box problem" of AI tools[20] would prevent the attorney from

being able to understand how this decision was reached. Perhaps in the future, when AI is a more

---

[18] Mata v. Avianca, 678 F. Supp. 3d 443, 448 (N.Y.S.D. 2023).

[19] Nate Raymond, *Ex-Trump Fixer Michael Cohen Says AI Created Fake Cases in Court Filing,* REUTERS, Dec. 29, 2023, https://www.reuters.com/legal/ex-trump-fixer-michael-cohen-says-ai-created-fake-cases-court-filing-2023-12-29/.

[20] The "black box problem" is a technology term that refers to a difficulty in being able to understand how an algorithm arrives at its conclusion. In AI, this problem is particularly prevalent, as developers generally still do not understand *how* AI systems make decisions. *See* Lou Blouin, *AI's Mysterious 'Black Box' Problem Explained,* U. MICH. NEWS, Mar. 6, 2023. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained.

reliable source, this tool will be helpful for juror research. But for now, while systems are generating fake information and it would be unclear who exactly would be accountable for the use of this misinformation, the danger of generating false information is too great.

     c.   Public Trust and Privacy Concerns

Another disadvantage AI tools pose is their ability to undermine public trust in the legal system. Given the large publicity around new AI tools, including their failings, it is no surprise that AI has not gained the trust of the general public. A recent global study revealed that the majority of people feel either ambivalent or distrustful of AI, and while most people surveyed agreed that AI use has a wide range of benefits, only about half agreed that those benefits outweighed the risks that AI poses.[21]

There are many reasons for this distrust,[22] including the bias and misinformation described above, but one of the biggest drivers of public distrust is concern for privacy. AI tools are able to collect more data on individuals than ever before, gathering everything from public social media postings to sensitive data, such as personal finances or healthcare information.[23] There is increasing concern that these tools will be able to collect sensitive data without the knowledge or consent of the individual.[24] This threat leads to a real public distrust of AI tools, as people become more and more aware of just how much of their data AI tools can access.

The current public distrust of AI tools could lead to decreased trust in the legal system. Public trust in the legal system is already low. A 2024 Gallup poll revealed that within the last

---

[21] Nicole Gillespie, et. al., *Trust in Artificial Intelligence: A Global Study*, U. QUEENSL. & KPMG AUSTL., 2023, https://ai.uq.edu.au/project/trust-artificial-intelligence-global-study.

[22] See Bhaskar Chakravorti, *AI's Trust Problem,* HARV. BUS. REV., May 3, 2024, https://hbr.org/2024/05/ais-trust-problem.

[23] Alice Gomstyn and Alexandra Jonker, *Exploring Privacy Issues in the Age of AI,* IBM, Sep. 30, 2024, https://www.ibm.com/think/insights/ai-privacy.

[24] *Id.*

year, public confidence in the judiciary has dropped to a record low of 35%.[25] If the courts were to institute AI tools now, at a time when AI still has major flaws to work out and the public can barely trust the legal system as it currently stands, the risk of negative effects of public trust falling even more is too high to risk.

IV.   Conclusion

It is undeniable that the recent case of Miguel Angel Peña-Rodriguez points out the ever-present risk of bias in juries and the need for new solutions to improve jury selection. While AI may be the solution to this problem someday, the risks that this technology poses in its current state are just too high to consider using AI at this time. It is hopeful that one day, developers will be able to further mitigate these risks and increase public trust in AI tools, but until that day comes, AI should not play a role in jury selection.

---

[25] Benedict Vigers and Lydia Saad, *Americans Pass Judgement on Their Courts,* GALLUP, Dec. 17, 2024, https://news.gallup.com/poll/653897/americans-pass-judgment-courts.aspx.