VOLUME 24, ISSUE 2, PAGES 31–46 (2023)

Criminology, Criminal Justice, Law & Society

CRIMINOLOGY, CRIMINAL JUSTICE, LAW & SOCIETY

E-ISSN 2332-886X
Available online at
https://scholasticahq.com/criminology-criminal-justice-law-society/

Content Analysis of Judges' Sentiments Toward Artificial Intelligence Risk Assessment Tools

Anna Fine,[a] Stephanie Le,[a] Monica K. Miller[a]

[a] *University of Nevada, Reno*

ABSTRACT AND ARTICLE INFORMATION

Artificial intelligence (AI) uses computer programming to make predictions (e.g., bail decisions) and has the potential to benefit the justice system (e.g., save time and reduce bias). This secondary data analysis assessed 381 judges' responses to the question, "Do you feel that artificial intelligence (using computer programs and algorithms) holds promise to remove bias from bail and sentencing decisions?" The authors created apriori themes based on the literature, which included judges' algorithm aversion and appreciation, locus of control, procedural justice, and legitimacy. Results suggest that judges experience algorithm aversion, have significant concerns about bias being exacerbated by AI, and worry about being replaced by computers. Judges believe that AI has the potential to inform their decisions about bail and sentencing; however, it must be empirically tested and follow guidelines. Using the data gathered about judges' sentiments toward AI, we discuss the integration of AI into the legal system and future research.

Article History:

Received March 5, 2023
Received in revised form July 4, 2023
Accepted July 5, 2023

Keywords:

artificial intelligence, judges, decision-making tools, community sentiment, algorithm aversion, algorithm appreciation, procedural justice, legitimacy

Corresponding author: Anna Fine, Interdisciplinary Social Psychology Ph.D. Program, Mailstop 1300 1664 N. Virginia St., Reno, NV, 89557, USA.
Email: afine@nevada.unr.edu

Artificial intelligence (AI) uses computer science and existing data to complete new problem-solving tasks (IBM Cloud Education, 2020). Specifically, AI uses algorithms created using large datasets to classify, analyze, and make predictions (Shroff, 2019). Algorithms have a set of automated instructions that are executed after initiation (Scott, 2021).

AI is a technological phenomenon with social, cultural, and political implications. Algorithms enhance many day-to-day tasks related to email communications, social media, web searches, and shopping services (Victor, 2021). They can also save lives. For example, Facebook implemented AI to detect suicidal posts (Cassata, 2019). Although AI potentially benefits society on a massive scale, a social-psychological perspective is necessary to understand human-AI interactions (Lindgren & Holmström, 2020). The social dimension is relevant to this research in that sentiments toward AI technology are socially constructed and come with hopes and fears that influence behaviors about interacting with this technology (Pinch & Bijker, 1984).

AI is already being developed and used in various areas of the justice system (Rigano, 2019), including decision-making tools used within the courts. This technology detects patterns that are difficult for humans to perceive; for instance, it predicts defendant recidivism risk and helps administrators distribute resources more effectively using predictive modeling (Henman, 2020). AI can extract information from lengthy legal documents, which can help decrease the tedious work of court personnel (Zadgaonkar & Agrawal, 2021). Police use this technology, too; it predicts areas with high crime rates and uses the information to allocate resources (Perry, 2013). Government bodies such as U.S. Homeland Security use it for automated facial recognition to identify potentially dangerous people on "watch lists" (Ritchie et al., 2021). These are a few examples of the application of AI technology in the justice system; however, the most common use of AI is algorithmic risk assessments used in bail and sentencing (McKay, 2020).

AI offers promise to reduce bias within the justice system because it bases its decisions on data rather than subjective perceptions of the legal decision-maker. However, this is entirely dependent on the nature and quality of the data. The complexity of AI in the justice system brings forth concerns of algorithmic bias such that the data might reflect bias (e.g., arrest histories for drug crimes by neighborhood, which is significantly influenced by race and socio-economic status).

AI is still relatively new, and the US government is working to introduce legislation to regulate it (Lee & Lai, 2022). There are no current federal regulations, although some states have introduced statewide legislation (National Conference of State Legislatures, 2022). For example, California has a bill pending (A 331) that requires creators and users of any automated decision-making tool to conduct a review of its impact. Specifically, why the software is being used, the expected positive outcomes, as well as how and where it will be used. This review will be submitted to the Civil Rights Department. Further, if these rules are not followed, then the software creators or users can be sued. Washington also has a bill pending (S 5356) that establishes guidelines for the obtaining and use of automated decision-making tools to protect the public and improve transparency. These are just two of the few states that have pending legislation in regard to automated decision-making tools (National Conference of State Legislatures, 2022).

Therefore, researchers need to understand judges' sentiments toward using AI courtroom technology, given that they will decide whether to adopt this technology. This secondary data content analysis explores judges' sentiments about algorithm aversion and appreciation, locus of control, procedural justice, and legitimacy. More specifically, this research evaluates whether judges believe that AI offers promise in reducing bias within bail and sentencing decisions, informing future research exploring judicial beliefs toward AI within the court setting.

Literature Review

Risk Assessment Tools in Courts

Judges use risk assessment tools in bail and sentencing. The purpose of a risk assessment tool is to help judges quickly classify pertinent information related to a defendant to determine the bail eligibility and the appropriate sentence. While judges are generally bound by law through sentencing guidelines, they also have discretion to mandate a sentence within these guidelines (Bushway & Piehl, 2001). Two commonly used risk assessment methods to assess recidivism are clinical and actuarial (Buskey & Woods, 2018). These methods differ in their ability to accurately predict recidivism; as such, the choice of method has implications for bail and sentencing decisions.
text.

Clinical versus Actuarial Models

Risk assessments use clinical and actuarial methods to predict a defendant's likely recidivism risk. Using the clinical method, professionals (e.g., forensic psychologists and clinicians) use their

personal experience, expertise, and intuition to differentiate between low and high-risk offenders (Mossman, 1994). The alternative method, actuarial, uses statistical instruments, algorithms, or AI techniques to predict levels of recidivism, which is a very new development with these assessments (Barabas et al., 2018).

One of the primary reasons why AI models have a high accuracy rate is their ability to process vast amounts of data quickly and accurately (Helm et al., 2020). They are able to identify patterns, analyze large amounts of data, and make predictions based on that data. Additionally, AI models learn and improve using machine learning techniques, which increase their accuracy over time (Helm et al., 2020). The accuracy of AI refers to the degree to which an AI system's predictions match the actual outcomes (Rueda et al., 2022). AI accuracy is essential for evaluating the performance and effectiveness of AI. AI has the ability to outperform human abilities and has higher predictive validity than human forecasters (Dietvorst et al., 2015).

AI is more accurate when making forecasts under uncertainty than human decision-makers (Dawes, 1979; Grove et al., 2000). Further, experienced forecasters who rely more heavily on a human's advice than an algorithm have lower accuracy (Logg et al., 2019). Even so, people are hesitant to trust AI and would prefer a human decision-maker, even when they are aware of the increased accuracy of AI (Diab et al., 2011; Eastwood et al., 2012). While the accuracy of the AI tool is important (Hoff & Bashir, 2015), fairness is also critical to trust in AI (Knowles et al., 2022).

Actuarial methods using AI technology hold promise as an effective way to help reduce bias in bail and sentencing; however, AI is only as good as its trained data, and there have been years of systemic racial bias within bail (Demuth & Steffensmeier, 2004b; Schlesinger, 2005; Turner & Johnson, 2005) and sentencing decisions (Demuth & Steffensmeier, 2004a; Spohn & Holleran, 2000; Western, 2006).

Therefore, it is essential to understand the limitations of AI. For example, Pro Publica assessed the COMPAS criminal algorithmic risk assessment for racial discrimination and found that Black defendants were twice as likely to be marked as high-risk than White defendants (Angwin et al., 2016). Moreover, White defendants were twice as likely to be classified as low-risk than Black defendants. Therefore, judges must be aware of the potential for AI to perpetuate bias if the algorithm is trained on racially biased data. While this assessment brought light to the dynamic and unclear role and effect of AI in risk assessments, their methodologies have been criticized. Jones (2020) used mathematical modeling to analyze the COMPAS data and found that Black individuals' risk rates were accurately calculated within 2% of actual rates, while White and Hispanic individuals were predicted to recommit crimes at a lower rate than they actually do.

While there are rules of moral philosophy and legal ethics for judges, there needs to be specific ethical guidelines for AI. There are generally no ethical guidelines for judges to follow specifically in regard to the tool's transparency, as well as fairness and accuracy, nor is there any law or rule that requires them to adhere to the tool's recommendation. Thus, beliefs (and statistics) about accuracy and fairness could guide judges' decisions on whether to use AI. These criteria are likely not the only ones that judges use to form their decisions and perceptions. Social psychological theory related to community sentiment can help researchers understand judges' perceptions of AI.

Community Sentiment of Artificial Intelligence

Community sentiment refers to a group of people's shared beliefs, attitudes, and opinions on a particular topic or issue (Miller & Chamberlain, 2015). It is influenced by various social, cultural, and environmental factors and significantly impacts the legal system and laws. Community sentiment often plays a role in shaping legislation (Burstein, 2003) and prioritizing legal issues that are of particular significance to the public (Burstein, 2006). When it comes to a specific policy question, community sentiment might be based on what the general public thinks, or it could only take into account the opinions of relevant experts and insiders. Changes in policy could be linked to community feelings, guided by the public's priorities for making policies (Miller & Chamberlain, 2015). Understanding community sentiment can help determine how judges perceive and utilize AI tools in the justice system.

Judges are the legal authority and have a consequential impact on the court system; therefore, it is essential to understand their community sentiment. Judges with negative beliefs towards AI might be less likely to use it or override the AI's decision. Conversely, judges with positive beliefs toward AI might be more willing to incorporate AI into their decision-making processes.

Developing and implementing successful AI tools (i.e., that have positive outcomes such as reducing bias but have few negative outcomes) in the justice system requires understanding community sentiment. By considering the beliefs of stakeholders, such as judges, AI tools can be designed and implemented in ways that are more likely to be accepted and utilized. This, in turn, can lead to more developing and testing—and ultimately promote better decision-making in the criminal justice system. Some

aspects of community sentiment include algorithm aversion and appreciation, locus of control, and perceptions of procedural justice and legitimacy.

Algorithm Aversion and Appreciation

Community sentiment about AI can be partially explained by algorithm aversion and algorithm appreciation. Dietvorst and colleagues (2015) coined the term algorithm aversion to describe a circumstance when people are unwilling or opposed to using an algorithm instead of a human, even when the algorithm has higher accuracy. However, in a series of studies, Logg and colleagues (2019) found that there are indeed circumstances (i.e., numeric estimates, forecasts on song popularity, romantic attraction) in which users experience algorithmic appreciation, which occurs when people prefer to use an algorithm over a human forecaster. These findings suggest that factors other than accuracy influence people's decision to trust AI.

Not only does community sentiment relate to reliance on AI, but it also depends on the level of uncertainty in the type of decision. In areas where there is no unavoidable randomness or unpredictability, like when doing arithmetic problems or remembering known facts, the best decision-making method can always give a perfect, error-free answer (Dietvorst & Bharti, 2020). However, errors occur even using the best decision-making tools in decision domains with unavoidable uncertainty (e.g., self-driving cars, criminal risk assessments, and medical diagnosing tools). Therefore, judges might be less likely to trust AI over their judgment due to the uncertain nature of criminal justice decisions. Locus of control offers insight into how uncertainty can influence trust in a decision-maker.

Locus of Control

Locus of control refers to a person's belief about how much they control their own lives and the events that affect them (Rotter, 1966). It is a concept that describes the degree to which people believe that they have power over the outcomes of their actions or whether they are primarily influenced by external factors such as luck, fate, or the actions of others (Rotter et al., 1972). Locus of control can significantly impact a person's attitude, behavior, and well-being (Wallston & Wallston, 1978).

People with a more internal orientation believe that outcomes are contingent on their behaviors (Sherman, 1973) and tend to trust their judgment over AI or other human decision-makers (Sharan & Romano, 2020). People who emphasize autonomy, personal achievement, individual responsibility, and competence might have a higher internal locus of control. People with a more external orientation believe that outcomes are contingent on outside forces (e.g., environment, luck, others) and are more open to outside forces having input regarding their decisions (Sherman, 1973).

Judges' locus of control orientation could impact their sentiment toward and willingness to rely on AI to reduce bias in bail and sentencing decisions. Judges might fall on one side of the scale, which could be related to their career choice. High-status occupations are associated with an internal locus of control (Smith et al., 1997). Given the prestige of a judicial career, judges might naturally lean toward having an internal locus of control, making them less inclined to view AI as a suitable decision-making tool over their judgment. Judges' sentiment might also relate to their concern about how AI could impact the perceived fairness of judicial procedures, as discussed next.

Procedural Justice

Procedural justice is the perceived fairness of decisions (Lemons & Jones, 2001) that results from the process by which decisions are made (Leventhal, 1980). In the legal system, procedural justice is critical to ensuring that the rights of all parties are protected and that decisions are impartial and transparent. If a decision-making process is fair, even unfavorable outcomes are considered procedurally just (Thibaut & Walker, 1975).

One of the most critical factors of procedural justice is that those affected by the decision have a voice (Lind & Tyler, 1988). Regarding AI and procedural justice, one of the key issues is ensuring that those affected by AI decision-making have the opportunity to question the decision-making process. Given its black-box nature, the process in which AI makes decisions is difficult to question. Defendants might not get to express their "voice," as an AI tool relies on statistics and does not allow the defendant to tell their side of the story. This lack of transparency can lead to perceptions of unfairness and undermine trust in the decision-making process. In order to uphold procedural fairness, it is necessary for the tools used to be transparent and explainable. Moreover, AI systems must have a procedure allowing avenues of appeal if legal actors deem the decision unfair. Within the legal system, AI can be used to aid judicial decision-making. However, AI must not undermine procedural justice.

Judicial use of AI tools could potentially undermine the perceived fairness of decisions.
For example, in the employment domain, employees perceive AI-based job interviews as less procedurally just than human-based interviews (Parasuraman et al., 2000). Regardless of accuracy, people tend to be less trusting of AI in various domains and contexts.

However, research suggests that including human elements in decision-making could increase procedural fairness (Lee et al., 2019). Therefore, ensuring the justice system implements AI systems to work alongside human decision-makers rather than replacing them is vital. Failure to promote procedural justice could lead to a lack of perceived legitimacy in the decision-making process, undermining community sentiment and negatively affecting social cohesion and cooperation. With the use of AI tools within the courtroom, the public might perceive the process as unfair or have skepticism towards a decision made by an AI tool. Social cohesion relies on the shared belief in fairness and effectiveness within the justice system. Therefore, if the decision-making process is viewed as unfair, then that could lead to a lack of social cohesion.

Legitimacy

Legitimacy is the belief that authorities, organizations, and social agreements are proper and impartial (Tyler, 2006), and retaining it relies upon the authorities adhering to procedural fairness norms (Clawson et al., 2001; Farnsworth, 2003). Governments are perceived as legitimate when they align with the group's norms, beliefs, and values (Zelditch, 2006) and their citizens regard them as deserving of support (Gurr, 1974). The judicial system needs to have a high level of legitimacy to maintain the rule of law, as legitimacy is its principal political capital (Gibson, 2006), and its absence means a lack of power (Zelditch, 2006).

For governments to appear legitimate, decisions must be fair, and everyone must have an opportunity to benefit at some point (Bühlmann & Kunz, 2011). The judiciary needs to have institutional legitimacy, which is the concept that the public will support their decisions and authority (Audette & Weaver, 2015) regardless of the satisfaction level of the decisions (Bühlmann & Kunz, 2011). Without legitimacy, the courts do not have enough power to rule against public opinion when necessary. Perceived bias in the judiciary can decrease the legitimacy of the courts (Ramirez, 2008), and therefore, judges have the incentive to maintain it.

The perception of judicial legitimacy could diminish when AI is involved in decision-making. Three relevant challenges for the perceived legitimacy of AI systems include 1) input legitimacy, which refers to citizens' disbelief that they have control over the data that are collected; 2) throughput legitimacy, which refers to citizens' lack of understanding of the procedures and the black box nature of AI; and 3) output legitimacy, which refers to citizens' doubt that AI can make more accurate decisions than humans (Starke & Lünich, 2020). Transparency, democracy, and accountability are all necessary for legitimacy. As applied to AI, the judiciary must use trustworthy AI regulated through guidelines and independent oversight (P. Andrews, 2022).

Maintaining procedural justice and legitimacy is essential to the success of the justice system and the judiciary. Therefore, it is crucial to understand how AI technology within the justice system influences the perception of legitimacy and procedural justice. AI technology is starting to show promise for reducing bias in bail and sentencing decisions. However, these tools do not come without pitfalls. For them to serve the created purposes, they need support from users. It is essential to understand users' sentiments toward their implementation. As there is a dearth of research exploring judges' sentiments toward AI, a qualitative approach allows for understanding some of their most pressing concerns.

Overview of Study

There is minimal research exploring judges' sentiments toward AI use in bail and sentencing decisions. Although AI tools are more accurate than human forecasters (Dietvorst et al., 2015), these tools still need to be improved, as they can reflect current systemic bias within the justice system. However, as tools improve over time and errors are reduced, it would be beneficial and efficient for judges to learn how to work with AI tools to inform their decisions. This study investigates judges' community sentiment toward AI used in bail and sentencing and whether AI holds promise in minimizing bias.

This research addresses general research questions:

1. Do judges have a generally negative or positive community sentiment regarding the use of AI?
2. Do judges' responses reflect an internal or external locus of control?
3. Do judges believe that the use of AI will produce a threat to procedural justice and legitimacy?
4. What are some benefits and concerns that judges express about AI as a decision-making tool?

Method

Participants and Procedure

These data were collected by The National Judicial College (NJC) in Reno, Nevada. In the January 2020 Question of the Month, judges were asked, "*Do you feel that artificial intelligence (using computer programs and algorithms) holds promise to*

remove bias from bail and sentencing decisions?" Responses came in from 381 judges who are alumni or have taken a course with The NJC in Reno, Nevada, who responded to the study. Judges responded "Yes" or "No" with a space to elaborate, which 168 judges did.[1]

Data Coding

We conducted a conceptual content analysis that involved analyzing the content of comments from judges who responded and elaborated on the survey question. The authors created an apriori coding scheme that captured the key themes within judges' comments based on the literature. Codes reflected the research questions and common themes that emerged from open-ended responses. The units of analysis were statements, and we separated judges' comments (e.g., 168) into statements (e.g., 325) that each encompassed one idea. For example, we split a judge's comment reflecting three ideas into three separate statements (e.g., if they mentioned three separate concerns about AI). This allowed us to measure how many times any judge mentioned a theme. Each open-ended response was analyzed to identify the presence of each theme.

For the general sentiment theme, the coders used 0 when the theme was negative, 1 when the theme was neutral, and 2 when the theme was positive. The themes within algorithm aversion and algorithm appreciation (e.g., utility, ethics, and willingness to learn) received a code of 0 (theme absent), 1 (algorithm aversion), and 2 (algorithm appreciation). For example, if the judge mentioned a utility concern, the code would reflect a 1. On the contrary, if the judge mentioned a utility benefit, the code would reflect a 2. The rest of the themes received a code of 0 (theme absent) or 1 (theme present) for each theme. Codes are not mutually exclusive, meaning the authors assessed each statement to determine if it reflected a particular theme. For example, in the statement, "It could be helpful to save time," the general sentiment theme was coded as 2 (positive) and the theme utility benefit would be a 1.

The first and second authors practiced coding together on 20 responses and then coded 40 responses alone. The first and second authors then divided the remaining statements for coding. The authors achieved a .89 Holsti's interrater reliability indicating that the coders perceived the codings exactly the same 89% of the time. This is typically considered an acceptable rate (Belur et al., 2021). Appendix A includes each theme and its definitions.

Results

RQ1: *Do Judges Have a Generally Negative or Positive Community Sentiment?*

In answering the yes/no question, most judges (n = 243; 64%) answered "no," that they did not believe AI was promising in removing bias from bail and sentencing decisions. In their write-in responses, judges had an overall negative sentiment toward using AI for bail and sentencing, as 229 statements were negative, 33 were neutral, and 63 were positive. Sentiment was often reflected as algorithm appreciation or aversion, which are two separate categories in the Community Sentiment Theme (see Appendix). Judges reflected algorithm appreciation regarding the tool's usefulness, as 48 statements mentioned some form of utility benefit (e.g., "providing data and patterns would be helpful") and a willingness to learn. Some judges experienced algorithm aversion, especially concerning the ethical disadvantages of using AI in bail and sentencing decisions (n = 15; e.g., "Because minorities are overrepresented in past arrests and criminal history given the reality of past justice system practice, we will be in a "Bias In/Bias Out" scenario for at least a couple generations more"). Further, some judges mentioned that these tools would not be useful (n = 10; e.g., "it cannot take into account the subtle aspects of human interaction, which is more telling than any paper test").

RQ2: *Do Judges' Responses Reflect an Internal or External Locus of Control?*

Judges' statements reflected a high internal locus of control. The idea of using AI for bail and sentencing decisions seemed to make judges experience a lack of control of the AI process (n = 12; e.g., "This process is hidden and always changing, which runs the risk of limiting a judge's ability to render a fully informed decision and defense counsel's ability to zealously defend their clients"). Overall, judges had high levels of uncertainty (n = 30; e.g., "With these facts, or lack thereof, how does a judge weigh the validity of a risk-assessment tool if he/she cannot understand its decision-making process?") compared to low levels of uncertainty (n = 10; e.g., "Based on the latest Malcolm Gladwell book I do"). Even more so, judges believed that AI threatened their autonomy and highly advocated that judicial discretion was the most important aspect of bail and sentencing decisions (n = 61; e.g., "We need independent judicial discretion.").

RQ3: *Do Judges Believe that the Use of AI will Produce a Threat to Procedural Justice and Legitimacy?*

Judges had major concerns about how using AI might threaten procedural justice, for instance, because defendants should be treated fairly and deserve to know that the decisions have been made with careful consideration (*n* = 31; e.g., "AI is not now nor will ever be able to deliver justice or mercy"). Further, many judges agreed that sentencing and fairness demand judges to consider the whole story and not just what is programmed in the code, reflecting the notion of "voice," which is an important part of procedural justice (*n* = 39; e.g., "We're sentencing people, not machines. Every case is different"). Similarly, judges also suggested that a procedure requires consideration of qualitative factors that only humans can comprehend (*n* = 84; e.g., "It will be another tool, but it cannot replace the human element, which must include the victim impact. AI can't replace this").

Concerning legitimacy, 16 statements asserted that the public might question the system's authority if the judiciary delegated their decisions to computers (e.g., "Having qualified judges with knowledge as to the law and people is crucial."). Further, 19 statements mentioned that, if AI is going to be used, then there need to be guidelines in place to ensure legitimacy (e.g., "Currently, there is no federal law that sets standards or requires the inspection of these tools, the way the FDA does with new drugs."). Judges value perceived legitimacy and believe that AI could potentially threaten it.

RQ4: *What are Some Benefits and Concerns Judges Express about AI as a Decision-making Tool?*

In addition to the utility benefits mentioned in RQ1, judges expressed some other benefits related to using AI in bail and sentencing. Judges' statements expressed an interest and willingness to learn more about developing technology to reduce bias in the courtroom (*n* = 11; e.g., "It should be studied. Empirical data will be necessary to evaluate the benefits and role it should play in bail consideration"). Others recognized their own potential for implicit bias and explained that AI might help reduce bias (*n* = 18; e.g., "Implicit bias is so hard to navigate through, and this may take some of it from the decision").

Judges' statements expressed many more concerns than benefits related to AI in the courtroom. A large number of statements expressed distress about the integrity of the data and programming (*n* = 73; e.g., "The problem is that the creators of the software are also human, which leaves open the chance that we create a biased system which has the additional harmful imprimatur of being 'bias free' and therefore seemingly unchallengeable"). Many were concerned about being replaced (*n* = 30; e.g., "What a dystopian nightmare awaits us when judges are replaced by so-called artificial intelligence!"). Further, some judges' statements explained that using AI would make bias worse (*n* = 26; "Computer programs may remove human biases, but they will only create new ones"). Other judges mentioned that AI lacks objective reasoning (*n* = 12; "I think that a machine would be less likely to question the data or its completeness").

Discussion

This study explored judges' community sentiment towards AI used in bail and sentencing and whether AI holds promise in addressing bias. Overall, this study demonstrated that judges reflect an internal locus of control, algorithm aversion, and have procedural justice and legitimacy concerns related to using AI to make decisions. If AI techniques are to be developed and implemented in ways that minimize negative aspects and maximize positive aspects of AI, it will be essential to gain judges' support. Understanding judges' concerns are the first step. Psychological theories can help guide the development of AI practices and the education of judges. Specifically, jurisdictions that want to develop just AI-related procedures should focus on the specific concerns of judges in this study, such as concerns about giving defendants and victims a "voice," which is a component of procedural justice. A judge could consider both the AI recommendation and the voices of the defendant and victims. Educating judges about these safeguards might reduce their fears, making them more willing to support the development and use of just AI.

The first major finding was that a majority of judges (64%) did *not* believe that AI would remove bias from bail and sentencing decisions and had an overall negative sentiment toward AI. However, some judges (36%) believed that AI *could* reduce bias. Judges reflected more algorithm aversion compared to algorithm appreciation, as they were generally concerned about the disadvantages; however, many did admit that the tool seemed generally helpful. This indicates that most judges are not yet convinced that AI has a place in the courtroom and that AI advocates have their work cut out for them.

Some judges' comments reflected a high internal locus of control, as they expressed that using AI as a decision-making tool made them feel a lack of control, a high level of uncertainty, and a threat to their judicial discretion. This aligns with previous research on locus of control and decision-making tools such that people with a more internal orientation tend to

trust their judgment over AI or other human decision-makers (Sharan & Romano, 2020). Thus, it is likely that only judges with low internal locus of control would be willing to adopt AI, even if it is developed in ways that maximize benefits (e.g., reduce bias).

Judges expressed that AI posed a threat to procedural justice and legitimacy. Some judges made clear that there would be no justice without the human element. Further, they had significant concerns about treating humans as statistics rather than people. Transparency is one way to promote procedural fairness (Lee et al., 2019) and legitimacy (de Fine Licht & de Fine Licht, 2020). Further, decision-makers who solely rely on AI are perceived to have low legitimacy (Starke & Lünich, 2020). Judges should not exclusively rely on AI to make judicial decisions but should be using it as a helpful tool to guide their decisions. Both judges who believed that AI held promise and those who did not agree that these decision-making tools should be *supplemental* tools and not a *replacement* for judges. These concerns should be taken into consideration in both the development of AI and training about AI.

Implications

This research can inform future interventions and continuing education courses for legal actors. Within the near future, judges will be exposed to a range of AI technology, from decision-making tools to evidence that appears in court. Therefore, education is necessary to integrate these tools into the justice system successfully. To help inform legal actors about this technology, researchers should study how judges learn to trust and interact with this technology. This research offers insight into judges' concerns, which should be considered when developing this technology. Judges feel a large amount of uncertainty due to the lack of transparency of AI. Future interventions should design training for judges that explains the underlying process of AI and the factors that are taken into account when evaluating risk. Reduced uncertainty and increased transparency could help some judges be more comfortable with AI.

To ensure procedural justice and legitimacy, there must be standardized ethical guidelines for technology used within the courts. AI holds promise to help reduce bias in bail and sentencing but requires more advancement in terms of ethical guidelines. Fjeld and colleagues (2020) reviewed reports from multiple organizations that outline principles for AI. They found eight key themes: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values. Many judges in our sample expressed concerns that fit these themes. Thus, these principles can help guide the development and implementation of AI in the courts to ensure procedural justice and legitimacy.

Limitations and Future Directions

This study has multiple limitations due to its design, population, and characteristic (e.g., secondary data). Although this study's strength is in assessing sentiment of a hard-to-reach population, it has a number of limitations. The first limitation is that the survey was only one question with an option to write in comments, which helped encourage judges to complete the survey but limits knowledge that can be gained. The survey did not include demographic questions, so we cannot tell whether sentiment might differ by the judges' age, gender, or time on the bench, for example.

The one question asked judges about their opinions about both bail *and* sentencing. Judges might have more experience in one context over the other or could have written their response to address only one context. Unless the judge specified within their response, it is difficult to discern whether they were referring to bail, sentencing, or both. A more in-depth survey would provide richer results. Even so, results can be useful in understanding general sentiment toward AI.

A second limitation concerns sampling bias. The survey was sent to everyone on the NJC listserv, but judges with contact with the NJC might differ from judges who do not (and thus did not have the chance to do the survey). Further, judges who chose to do the survey might differ from those who did not. For instance, judges with strong opinions (for or against AI) might have been more likely to respond than those with weak opinions. Thus, our sample likely does not represent all judges. Importantly, many judges in this sample might not be familiar with AI, and the survey did not measure familiarity. Thus, it is impossible to tell whether the sentiment of judges who are familiar with AI differs from those who are not. Future studies should test the exposure hypothesis, which suggests that contact with a target (e.g., AI) can change one's attitudes (Zajonc, 2001). A broader sample would allow researchers to assess attitudes of judges with a greater range of attitudes and experiences.

The last major limitation is due to the nature of secondary data analysis. While it can be a valuable method for conducting research, it also has some limitations. First, we had no control over how the data were collected or the development of the question. Further, we could not control for factors that can influence results. For example, NJC collected this data in January 2020, before the COVID-19 pandemic. Given the global and grand scale of the pandemic, there is a possibility that judges' perspectives toward AI have changed. For example, many judges learned

to use Zoom. Such experience might make some judges more open to using AI, given the increased reliance on technology due to the pandemic. Moreover, a person's experience with this technology can predict their preference toward AI (Kramer et al., 2018). As mentioned above, the exposure effect could explain this phenomenon, which suggests that ongoing exposure to a stimulus increases the positive affect toward that stimulus (Zajonc, 2001).

Our study can help inform future research. First, due to the nature of the survey only having one question, future research should conduct a more extensive experimental survey to measure judges' experience, technology use, and acceptance, as well as other individual difference measures. Next, given the potential selection bias, future research should gain a representative sample of judges to ensure generalizable results. Ultimately, as public opinion can significantly impact policy-making and legal decision-making, it is important for future research to investigate how the *public* perceives the utilization of AI in the judicial system. Ultimately, as public opinion can significantly impact policy-making and legal decision-making, it is important for future research to investigate how the public perceives the utilization of AI in the judicial system.

Conclusion

Bail and sentencing decisions are highly consequential decisions and judges are rightly concerned that such decisions are fair and bias-free. Although AI has the potential to *minimize* some biases inherent in human decisions, AI can also *inject* bias into decisions if the data it relies on are biased. AI has other benefits such as speeding up decision processes and reducing caseloads and thus is attractive to many proponents. Although AI might be an attractive solution to many issues, there is no standardization or guidelines to instruct judges as to how to use these tools in their bail and sentencing decisions. The time is ripe to develop fair algorithms and standardized procedures and to gain the trust of judges who will use them. A good first step is to understand judges' sentiment toward AI. Overall, judges were fairly unsupportive of AI, indicating that most are not ready to implement AI, as they currently understand it. Judges listed a number of benefits and concerns about using AI in bail and sentencing decisions. These can be the basis for developing AI procedures that will maximize benefits and minimize these concerns. Education about AI can be tailored to alleviate these concerns for judges who might misunderstand AI or its procedures.

Given that uncertainty plays a prominent role in trusting AI technology, locus of control offers a unique perspective to guide how judges perceive the use of AI within their courtroom. Specifically, many judges experience an internal locus of control, as they see AI as a loss of autonomy and that they have less control over the decisions made within their court. Understanding how AI use within the courts influences perceived procedural justice principles and legitimacy is essential. Some courtrooms are implementing this technology, yet there is a dearth of research exploring its consequences for the legal system. Researchers must be proactive rather than reactive if they are to help develop and implement AI that will promote just outcomes.

References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.

Andrews, P. (2022, October 13). Designing for legitimacy. *Apolitical.* https://apolitical.co/solution-articles/en/designing-for-legitimacy

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. In K. Martin (Ed.), *Ethics of data and analytics* (pp. 254–264). Auerbach Publications.

Audette, A. P., & Weaver, C. L. (2015). Faith in the court: Religious out-groups and the perceived legitimacy of judicial decisions. *Law & Society Review*, *49*(4), 999–1022. https://doi.org/10.1111/lasr.12167

Barabas, C., Virza, M., Dinakar, K., Ito, J., & Zittrain, J. (2018, January). Interventions over predictions: Reframing the ethical debate for actuarial risk assessment. *Proceedings of Machine Learning Research*, *81*, 62–76. https://proceedings.mlr.press/v81/barabas18a.html

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Bühlmann, M., & Kunz, R. (2011). Confidence in the judiciary: Comparing the independence and legitimacy of judicial systems. *West European Politics*, *34*(2), 317–345. https://doi.org/10.1080/01402382.2011.546576

Burstein, P. (2003). The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, *56*(1), 29–40. https://doi.org/10.1177/106591290305600103

Burstein, P. (2006). Why estimates of the impact of public opinion on public policy are too high: Empirical and theoretical implications. *Social Forces*, *84*(4), 2273–2289. https://doi.org/10.1353/sof.2006.0083

Bushway, S. H., & Piehl, A. M. (2001). Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law & Society Review*, *35*(4), 733–764. https://doi.org/10.2307/3185415

Buskey, B., & Woods, A. (2018). *Making sense of pretrial risk assessments*. National Association of Defense Lawyers. https://www.nacdl.org/Article/June2018-MakingSenseofPretrialRiskAsses

Cassata, C. (2019, December 20). *Facebook using artificial intelligence to help suicidal people*. Healthline. https://www.healthline.com/health-news/facebook-artificial-intelligence-help-suicidal-people

Clawson, R. A., Kegler, E. R., & Waltenburg, E. N. (2001). The legitimacy-conferring authority of the US Supreme Court: An experimental design. *American Politics Research*, *29*(6), 566–591. https://doi.org/10.1177/1532673X01029006002

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582. http://dx.doi.org/10.1037/0003-066x.34.7.571

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society*, *35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w

Demuth, S., & Steffensmeier, D. (2004a). Ethnicity effects on sentence outcomes in large urban courts: Comparisons among White, Black, and Hispanic defendants. *Social Science Quarterly*, *85*(4), 994–1011. https://doi.org/10.1111/j.0038-4941.2004.00255.x

Demuth, S., & Steffensmeier, D. (2004b). The impact of gender and race-ethnicity in the pretrial release process. *Social Problems*, *51*(2), 222–242. https://doi.org/10.1525/sp.2004.51.2.222

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay perceptions of selection decision aids in US and non-US samples. *International Journal of Selection and Assessment*, *19*(2), 209–216. https://doi.org/10.1111/j.1468-2389.2011.00548.x

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology. General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward decision-making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458–468. https://doi.org/10.1002/bdm.741

Farnsworth, S. J. (2003). Congress and citizen discontent: Public evaluations of the membership and one's own representative. *American Politics Research*, *31*(1), 66–80. https://doi.org/10.1177/1532673X02238580

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI* (Research Publication No. 2020-1). Berkman Klein Center. https://dx.doi.org/10.2139/ssrn.3518482

Garrett, B., & Monahan, J. (2019). Assessing risk: The use of risk assessment in sentencing. *Judicature*, *103*(2), 6–16. https://judicature.duke.edu/articles/assessing-risk-the-use-of-risk-assessment-in-sentencing/

Gibson, J. L. (2006). Judicial institutions. In R. A. Rhodes, S. A. Binder, & B. A. Rockman (Eds.), *The Oxford handbook of political institutions* (pp. 514–534). Oxford University Press.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Gurr, T. (1974). Persistence and change in political systems, 1800–1971. *American Political Science Review*, *68*(4), 1482–1504. https://doi.org/10.2307/1959937

Harris, H. M., Gross, J., & Grumbs, A. (2019). *Pretrial risk assessment in California.* Public Policy Institute of California.

https://www.ppic.org/wp-content/uploads/pretrial-risk-assessment-in-california.pdf

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, *13*, 69–76.

Henman, P. (2020). Improving public services using artificial intelligence: Possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, *42*(4), 209–221. https://doi.org/10.1080/23276665.2020.1816188

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

IBM Cloud Education. (2020, June 3). *What is artificial intelligence?* https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

Jones, L. (2020, October 6). *ProPublica's misleading machine bias*. Medium. https://medium.com/@llewhinkes/propublicas-misleading-machine-bias-19c971549a18

Knowles, B., Richards, J. T., & Kroeger, F. (2022). *The many facets of trust in AI: Formalizing the relation between trust and fairness, accountability, and transparency*. arXiv. https://arxiv.org/abs/2208.00681

Kramer, M. F., Schaich Borg, J., Conitzer, V., & Sinnott-Armstrong, W. (2018, December). When do people want AI to make decisions? *Proceedings of the 2018 Aaai/Acm Conference On AI, Ethics, and Society*, 204–209. https://doi.org/10.1145/3278721.3278752

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–26. https://doi.org/10.1145/3359284

Lee, N. T., & Lai, S. (2022, May 17). The U.S. can improve its AI governance strategy by addressing online biases. *Brookings*. https://www.brookings.edu/blog/techtank/2022/05/17/the-u-s-can-improve-its-ai-governance-strategy-by-addressing-online-biases/

Lemons, M. A., & Jones, C. A. (2001). Procedural justice in promotion decisions: Using perceptions of fairness to build employee commitment. *Journal of Managerial Psychology*, *16*(4), 268–281. https://doi.org/10.1108/02683940110391517

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in Theory and Research* (pp. 27–55). Springer.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Lindgren, S., & Holmström, J. (2020). Social science perspective on artificial intelligence: Building blocks for a research agenda. *Journal of Digital Social Research*, *2*(3), 1–15. https://doi.org/10.33621/jdsr.v2i3.65

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

McKay, C. (2020). Predicting risk in criminal procedure: Actuarial tools, algorithms, AI and judicial decision-making. *Current Issues in Criminal Justice*, *32*(1), 22–39. https://doi.org/10.1080/10345329.2019.1658694

Miller, M. K., & Chamberlain, J. (2015). "There ought to be a law!": Understanding community sentiment. In M. K. Miller, J. A. Blumenthal, & J. Chamberlain (Eds.), *Handbook of community sentiment* (pp. 3–28). Springer. https://doi.org/10.1007/978-1-4939-1899-7_1

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology*, *12*(1), 489–513. https://doi.org/10.1146/annurev-clinpsy-021815-092945

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*(4), 783–792. https://doi.org/10.1037/0022-006X.62.4.783

National Conference of State Legislatures. (2022, January 5). *Legislation related to artificial intelligence*. https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics. Part A, Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Rand Corporation.

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 399–441. https://doi.org/10.1177/030631284014003004

Ramirez, M. D. (2008). Procedural perceptions and support for the U.S. Supreme Court. *Political Psychology*, *29*(5), 675–698. https://doi.org/10.1111/j.1467-9221.2008.00660.x

Rigano, C. (2019). Using artificial intelligence to address criminal justice needs. *National Institute of Justice Journal*, *280*, 1–10. https://www.ojp.gov/pdffiles1/nij/252038.pdf

Ritchie, K. L., Cartledge, C., Growns, B., Yan, A., Wang, Y., Guo, K., Kramer, R. S. S., Edmond, G., Martire, K. A., San Roque, M., & White, D. (2021). Public attitudes towards the use of automatic facial recognition technology in criminal justice systems around the world. *PloS One*, *16*(10), e0258241–e0258241. https://doi.org/10.1371/journal.pone.0258241

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, *80*(1), 1–28. https://doi.org/10.1037/h0092976

Rotter, J. B., Chance, J. E., & Phares, E. J. (1972). *Applications of a social learning theory of personality*. Holt, Rinehart, and Winston.

Rueda, J., Rodríguez, J. D., Jounou, I. P., Hortal-Carmona, J., Ausín, T., & Rodríguez-Arias, D. (2022). "Just" accuracy? Procedural fairness demands explainability in AI-based medical resource allocations. *AI & Society*, 1–12. https://doi.org/10.1007/s00146-022-01614-9

Schlesinger, T. (2005). Racial and ethnic disparity in pretrial criminal processing. *Justice Quarterly*, *22*(2), 170–192. https://doi.org/10.1080/07418820500088929

Scott, A. (2021). *Difference between algorithm and artificial intelligence*. Data Science Central. https://www.datasciencecentral.com/difference-between-algorithm-and-artificial-intelligence/

Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, *6*(8), e04572. https://doi.org/10.1016/j.heliyon.2020.e04572

Sherman, S. J. (1973). Internal-external control and its relationship to attitude change under different social influence techniques. *Journal of Personality and Social Psychology*, *26*(1), 23–29. https://doi.org/10.1037/h0034216

Shroff, R. (2019, September 25). *Artificial intelligence explained in simple terms*. Medium. https://medium.com/mytake/artificial-intelligence-explained-in-simple-english-part-1-2-1b28c1f762cf

Smith, P. B., Dugan, S., & Trompenaars, F. (1997). Locus of control and affectivity by gender and occupational status: A 14 nation study. *Sex Roles*, *36*(1–2), 51–77. https://doi.org/10.1007/BF02766238

Spohn, C., & Holleran, D. (2000). The imprisonment penalty paid by young, unemployed Black and Hispanic male offenders. *Criminology*, *38*(1), 281–306. https://doi.org/10.1111/j.1745-9125.2000.tb00891.x

Starke, C., & Lünich, M. (2020). Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, *2*(1). https://doi.org/10.1017/dap.2020.19

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. L. Erlbaum Associates.

Turner, K. B., & Johnson, J. B. (2005). A comparison of bail amounts for Hispanics, Whites, and African Americans: A single county analysis. *American Journal of Criminal Justice*, *30*(1), 35–53. https://doi.org/10.1007/BF02885880

Tyler, T. R. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, *57*(1), 375–400. https://doi.org/10.1146/annurev.psych.57.102904.190038

Victor, A. (2021, July 24). 10 uses of artificial intelligence in day to day life. *Daffodil*. https://insights.daffodilsw.com/blog/10-uses-of-artificial-intelligence-in-day-to-day-life

Wallston, B. S., & Wallston, K. A. (1978). Locus of control and health: A review of the literature. *Health Education Monographs*, *6*(1), 107–117. https://doi.org/10.1177/109019817800600102

Western, B. (2006). *Punishment and inequality in America*. Russell Sage Foundation.

Zadgaonkar, A. V., & Agrawal, A. J. (2021). An overview of information extraction techniques for legal document analysis and processing. *International Journal of Electrical and Computer Engineering*, *11*(6), 5450–5457. https://doi.org/10.11591/ijece.v11i6.pp5450-5457

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*, *10*(6), 224–228. https://doi.org/10.1111/1467-8721.00154

Zelditch, M., Jr. (2018). Legitimacy theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 340–371). Stanford University Press.

About the Authors

Anna Fine, M.S., graduated from Arizona State University with a Master of Science in Psychology. She is a PhD student in the Interdisciplinary Social Psychology Ph.D. Program at the University of Nevada, Reno. Her research interests include how judges and the general public perceive artificial intelligence and its implementation within society, especially within the court system. She would like to acknowledge both Shawn Marsh and Monica K. Miller for their consistent encouragement to follow my research interests. Correspondence concerning this article should be addressed to Anna Fine, Interdisciplinary Social Psychology Ph.D. Program, Mailstop 1300 1664 N. Virginia Street, Reno, NV 89557. Email: afine@nevada.unr.edu.

Stephanie Le is a first-year Master of Sociology student at the University of Nevada, Reno. Her bachelor's degree is in Education and is obtaining her Master's degree in Sociology in order to pivot out of direct instruction into the realm of global non-profit and non-government organizations related to Education. Her research interests involve non-profit and non-government organizations, Global Education pedagogy and curriculum design, social justice, and anti-racism. She would like to acknowledge and thank her two faculty advisors Clayton Peoples and Monica K. Miller for their continued advice and support.

Monica K. Miller, J.D., Ph.D., is a Foundation Professor at the University of Nevada, Reno. She has appointments in the Criminal Justice Department and the Interdisciplinary Social Psychology PhD program and teaches in the Judicial Studies program. Her research is at the cross-section of law and social sciences, with focus on theory-based research that promotes well-being in the real world.

Appendix

Table 1: Identified Themes, Categories within Each Theme, Their Definitions, the Number of Times That a Category Appeared in Their Response, and Whether They Agreed or Disagreed with the Category

Theme	Category	Operational Definition	Number of Times Appeared
Community Sentiment	Overall Comment Support	Measurement of how supportive the comment was towards AI.	325 Negative = 229; Neutral = 33; Positive = 63
Algorithm Aversion & Algorithm Appreciation	Utility	Whether or not the judge thinks the tool will be (e.g., I think it could help me make decisions).	58 Utility Benefit = 48; No utility benefit = 10
	Ethics	Ethical advantage or disadvantage (e.g., it would be wrong to treat people like statistics).	15 Ethical advantage = 0; Ethical disadvantage = 15
Locus of Control	Threat to Autonomy/Judicial Discretion	Judge's perceived threat of autonomy and discretion (e.g., Judges should make their own decisions).	61 Threat to autonomy = 61
	High Level of Uncertainty	Judges' level of uncertainty about AI algorithms and how they work. (e.g., I am unsure of how AI can help).	40 High uncertainty = 30
	Control	Judge's perceived amount of control in the process (e.g., The AI is a black box and there is no way to understand the process behind its decision).	12 No control = 12
Procedural Justice	Threat to Justice	Judges' perceptions of AI as a threat to justice (e.g., Further, rigid adherence to any system without a common sense review is problematic and can lead to very unjust results.).	31 Threat to Justice = 31
	Not One Size Fits All	Sentencing and fairness demand judges to consider the whole story of the —not just whatever is programmed in the code (e.g., humans are complicated and a code cannot judge that).	39 Not one size fits all = 39
	Human Element	Judges' perceptions about having human involvement (e.g., humans need to be	84 Yes = 84

		involved in the decision-making process)	
Legitimacy	Threat to perceived legitimacy of the system	How people might view the justice system if an AI is helping decisions rather than the judge making the decisions themselves (e.g., "Removing bias" sounds like "sentencing guidelines," which means sentencing decision authority is being taken from judges.).	16 Threat to legitimacy = 16
	Guidelines needed	Need strict programming/guidelines/ testing/oversight/checks/ audits/scientific research (e.g., there needs to be some sort of set guidelines).	19 Guidelines needed = 19
Benefit	Judge Bias	Judges' admittance of their own biases	18 Judge bias present = 18
	Willingness to Learn	Judge's willingness to learn about AI (e.g., I need to see research in this area).	10 Willing to learn = 10
Concern	Bias in Data	There is bias within the data based on previously biased decisions within the criminal justice system. (e.g., garbage in garbage out).	73 Bias in data = 73
	Replace Judges	This is in reference to judges' fear about being replaced by AI (e.g., how about full robot courts).	30 Replace judges = 30
	Make Bias Worse	Make implicit bias worse (e.g., My experience is that non-white defendants have longer criminal histories due to more police contact. A defendant's criminal history is a significant element of the algorithm, which produces disparate results).	26 Make bias worse = 26
	Lack of Objective Reasoning	AI's lack of objective reasoning (e.g., AI cannot take into account the subtle aspects of human interaction, which is more telling than any paper test).	12 Lack of objective reasoning = 12

Endnotes

[1] Given that this was a secondary data analysis, IRB approval was unnecessary. Parts of this data were described in a general, non-scientific way in a brief news article in the Judicial Edge Today newsletter, which The National Judicial College publishes. It can be found here: https://www.judges.org/news-and-info/judges-remain-skeptical-on-whether-artificial-intelligence-can-make-decisions-more-fairly-than-they-can