Discover Artificial Intelligence

Research

Judicial leadership matters (yet again): the association between judge and public trust for artificial intelligence in courts

Anna Fine[1] · Shawn Marsh[1,2,3]

Received: 21 February 2024 / Accepted: 7 June 2024
Published online: 28 June 2024

Abstract
Artificial intelligence (AI) is rapidly expanding in myriad industries and systems. This study sought to investigate public trust in using AI in the criminal court process. While previous research has identified factors that influence trust in AI, such as perceived accuracy and transparency of algorithms, less is known about the role of influential leaders—such as judges—in shaping public trust in new technology. This study examined the relationship between locus of control, anthropomorphism, cultural values, and perceived trust in AI. Participants completed a survey assessing their perceptions of trust in AI in determining bail eligibility, bail fines and fees, sentencing length, sentencing fines and fees, and writing legal documents (e.g., findings and disposition). Participants were more likely to trust AI performing financial calculations rather than determining bail eligibility, sentence length, or drafting legal documents. Participants' comfort with AI in decision-making also depended on their perceptions of judges' trust in AI, and they expressed concerns about AI perpetuating bias and the need for extensive testing to ensure accuracy. Interestingly, no significant association was found with other participant characteristics (e.g., locus of control, anthropomorphism, or cultural values). This study contributes to the literature by highlighting the role of judges as influential leaders in shaping public trust in AI and examining the influence of individual differences on trust in AI. The findings also help inform the development of recommended practices and ethical guidelines for the responsible use of AI in the courts.

Keywords Artificial intelligence · Law · Judges · Perceptions · Trust

Artificial intelligence (AI) uses computer programming, algorithms, and large data sets to perform tasks usually conducted by humans, such as decision-making, reasoning, and learning from past experiences [20]. Rapidly expanding implementation of AI increases the likelihood that it will be used to make potentially highly consequential decisions that affect public lives. For example, and as elaborated further below, some criminal justice system jurisdictions already use algorithmic risk assessments to help determine if one is eligible for bail and the length of a sentence upon conviction [10, 80]. As of 2022, at least 60 jurisdictions nationally use algorithmic risk assessments in bail decisions [15]. Multiple states also use algorithmic risk assessments to guide sentencing decisions (e.g., Arizona, Colorado, Virginia, Washington, etc.) [2]. It is important to note that AI is currently being implemented with limited regulation [53].

Multiple factors can influence public trust in AI, and while there is relatively little research on this topic, judges may play a key role. Judges can shape public opinion of AI as respected legal experts through their legal decisions and public statements. This study's theoretical framework suggests that public trust in AI can be significantly influenced by the

✉ Anna Fine, afine@nevada.unr.edu; Shawn Marsh, shawnm@unr.edu | [1]Interdisciplinary Social Psychology Ph.D. Program, University of Nevada, 1664 N. Virginia Street, Mailstop 1300, Reno, NV 89557, USA. [2]Judicial Studies Graduate Program, University of Nevada, Reno, USA. [3]Communication Studies, University of Nevada, Reno, USA.

perceived trust of judges, aligning with the broader literature on expert influence [65]. When judges rule on AI-related cases or make public statements about AI, they can impact public perception of the technology and its societal effects. Therefore, judges' trust in AI, like experts in other fields, can significantly shape public opinion and acceptance of AI.

Locus of control refers to whether someone attributes their outcomes to internal or external forces. Individuals with a high internal orientation trust that outcomes are derived from their behaviors, emphasizing personal achievement, autonomy, individual responsibility, and competence [74]. Individuals with a high external orientation believe that outcomes are based on outside forces such as luck, environment, or others and are more willing to receive outside advice. The extent to which a person has an internal or external orientation might relate to their perceptions of AI in the legal system.

A person's cultural perspective could be associated with their perceptions of AI. Hofstede's cultural dimensions, trust, and locus of control theory offer unique perspectives to explain judges' judgment and decision-making concerning adopting predictive algorithms within their court. Hofstede's cultural framework simplifies the multidimensional cultural construct, encompassing social norms, beliefs, and attitudes [45]. Looking at cultural dimensions can be extremely informative and beneficial to understanding a larger range of people worldwide.

Another factor influencing perceptions is how much an individual anthropomorphizes AI. Anthropomorphism is the extent to which people attribute human characteristics to nonhumans [92]. People might trust AI tools more if they anthropomorphize them more often. For example, drivers of automatic vehicles gained trust in the vehicle as it gained more anthropomorphic features [92]. This research aims to understand how participants' perceptions of judges' trust in AI influence their level of trust. Further, this study will expand on the literature by testing locus of control, anthropomorphism, and cultural values as predictors of trust in AI. Understanding these cultural and psychological factors provides a foundation for evaluating the methods used in risk assessment within the justice system.

1 Clinical vs. actuarial

Risk assessments for predicting recidivism use clinical or actuarial methods. Clinical methods involve experts such as forensic psychologists and clinicians who rely on their personal experience and intuition [62], while actuarial methods utilize statistical tools like algorithms and AI to determine a defendant's risk level [5]. Although there is no standardized approach to selecting a method, evidence suggests that actuarial methods are more effective than clinical methods [40].

The use of accurate methods in the justice system is critical due to the well-established disparities in bail [24, 84] and sentencing decisions [24, 25, 79, 90] associated with race. Actuarial methods are more accurate than clinical methods in health and human behavior studies [40]. Using such methods in bail settings could help minimize the impact of race on these decisions. Therefore, understanding the accuracy of different decision-making methods in bail settings is crucial in addressing racial disparities in the justice system.

2 Accuracy

The accuracy of AI refers to the extent to which an AI system's predictions match the actual outcomes [47]. It is essential for evaluating the performance and effectiveness of AI. One of the key advantages of using AI models is their high accuracy rates due to their ability to quickly and accurately process vast amounts of data, identify patterns, and make predictions based on the data analyzed. AI models continually learn and improve through machine learning techniques, enhancing their accuracy over time [8].

Research has shown that AI has higher predictive validity than human forecasters and outperforms human abilities, particularly in making forecasts under uncertainty [22, 27, 40]. Experienced forecasters who rely more heavily on human advice than an algorithm tend to have lower accuracy [54]. Despite these findings, individuals still resist trusting algorithms and prefer human decision-makers, even when aware of the algorithm's increased accuracy [27].

Experts underestimate the effects of cognitive biases and motivated reasoning on their evaluations. Cognitive biases influence human evaluations in bail settings [36, 52] and mold expert recidivism predictions. Actuarial assessments, on the other hand, tend to be more precise as they reduce human bias in the forecasting process [39, 43, 60]. Nevertheless, some professionals believe their evaluations surpass others [19, 28], a notion connected to the *bias blind spot* [67]. A recent study of forensic specialists illustrated this, as participants viewed themselves as less biased than their peers [63]. Given the potential advantages of using AI in bail and sentencing settings, it is crucial to understand the utilization of AI in these contexts. It is just as important to note the limitations and challenges of AI in these contexts, such as algorithmic bias against marginalized groups [55]. Next, we will review recent research on using AI for risk assessment in the criminal justice system and examine this approach's potential benefits and limitations.

3 Bail and sentencing

Across the United States, courts have implemented procedures for utilizing risk assessments, with over 60 jurisdictions employing them in arraignment hearings [10] and up to 20 states incorporating them into sentencing processes [80]. However, there is currently no standardized approach to determining which type of risk assessment to use, whether clinical or actuarial.

AI already shows immense promise at the government level [86] and has already been established in taxation [83], health and safety [61], Social Security benefits determinations [37], and the Securities and Exchange Commission [6]. AI has also shown promise in improving the justice system. Algorithmic risk assessments are used in pretrial detention [57] and sentencing [59], as well as probation and parole [49]. There have been multiple instances in recent years that show the promising potential of AI.

San Francisco is using AI to blind police reports so that prosecutors are unaware of the defendant's race [34]. In Kansas, officials use machine learning to divert low-risk defendants from jail to mental health services [7]. In another example, more than 60,000 marijuana convictions in California were made eligible to be cleared through Proposition 64 [18]. An algorithm identified and cleared all eligible cases using an automated process (i.e., no hearing, attorney, or petitions needed). Given this population's disproportionate number of Black individuals, the algorithm helped negate historical bias [18]. Finally, one research team found that machine learning can help improve legal outcomes. They built an algorithm that resulted in 25% fewer crimes committed while defendants were released on bail, 45% lower capacity within the jails, and reduced overall racial disparities [48]. Therefore, AI offers promise to help increase the fairness of these consequential decisions.

While AI can benefit the justice system, it faces significant challenges and limitations, such as algorithmic bias, transparency, accountability, and privacy concerns. If training data contain biases, AI systems could perpetuate or exacerbate these biases, leading to racial discrimination and mass incarceration [55, 77]. The black-box nature of some AI algorithms makes it difficult to understand their decisions, resulting in lack of transparency and accountability [17, 26, 87]. Privacy and data protection concerns also arise, including informed consent, surveillance, and data rights violations [68, 88]. While the Fourth and Fifth Amendments offer some protections, there is limited case law on AI's intersection with privacy rights [31, 78]. Public perception of judiciary trust in AI is crucial for successful implementation in court decisions, making it important to understand and foster trust in AI's use within the justice system.

Public trust is essential for AI's acceptance and further progression and development [76]. Although a substantial body of literature explores interpersonal dynamics and trust (i.e., between humans), the literature on human–computer interaction and trust is limited. However, findings from relevant research suggest people are unsure of how much they should support and trust the use of AI in their day-to-day lives. For example, Americans express mixed support for the development of AI, believe it should be carefully managed, and have various levels of trust in different organizations to develop and oversee AI for the public's best interest [95]. Further, when asked about a list of AI-based products and services, over 40% of respondents said they would not trust one [50].

4 Trust in AI

The human–computer interaction literature has been exploring trust in AI. The extent to which someone trusts AI can vary by domain and its use but also by individual differences. Glikson and Woolley [38] discerned factors influencing emotional and cognitive trust in AI. Anthropomorphism predicted emotional trust, and transparency predicted cognitive trust. Trust plays an essential role in the intention to use AI, and there are likely differences when users decide to trust or distrust AI.

Algorithm aversion occurs when an individual trusts a human predictor over an algorithm [27]. People experience this type in highly uncertain domains (e.g., self-driving cars and medical diagnostics). In contrast, algorithm appreciation occurs when people prefer algorithms over human predictors in domains with less uncertainty (e.g., numeric estimates, forecasts on song popularity, and romantic attraction) [54]. Therefore, it is likely that people will be less likely to trust the use of AI in the criminal justice system, given the high level of uncertainty. Locus of control offers a potential explanation of how uncertainty might affect the level of trust in AI.

5 Locus of control

Locus of control refers to the subjective assessment between individual characteristics and outside circumstances encompassing experienced outcomes [70, 71]. Individuals range on a scale of orientation between internal and external. Given that people with an external locus of control are more open to outside forces, they might be more open to using AI technology within the courtroom than those with an internal locus of control orientation.

Although locus of control appears to stem from individual experiences, it is also plausible that the broader societal context may impact the LOC of various groups. Twenge et al. [85] conducted two meta-analyses examining the LOC in college students and children from 1960 to 2002. Both groups steadily shifted towards an external locus of control over time. Furthermore, the COVID-19 pandemic may have prompted a general shift in the population towards an external orientation, given the heightened feelings of helplessness and stress experienced during this period. One study found that the locus of control shifted from internal to external for social work professionals and college students [58]. Some AI systems have anthropomorphic qualities, which turn them into social actors [89]. Thus, it is important to explore how the extent to which people anthropomorphize AI influences their perceptions of AI.

6 Anthropomorphism

There is a distinction between the type of anthropomorphic cues that are attributed to embodied agents like robots (e.g., voice, body movements, facial expressions) and disembodied agents like chatbots (e.g., personality, name, gender, voice) [3]. How people classify people and objects influences how they process information regarding self-report and behavioral measures [21]. On the one hand, objects are generally evaluated based on their quality and utility. On the other hand, people are less likely to be evaluated on their functionality and rather on their interpersonal qualities (e.g., warmth). Thus, peoples' perceptions of anthropomorphized objects might be sensitive to information within the interpersonal realm.

Anthropomorphism has been found to influence the level of trust in AI. When consumers were primed to think about their car in anthropomorphic terms, they were less likely to replace it, gave less weight to functionality, and more weight in terms of interpersonal descriptors when making replacement decisions [12]. However, some studies have shown weak or no effects on anthropomorphism and trust [29, 41]. Just as perceived anthropomorphism can influence perceptions of AI, culture is another factor that might come into play as an individual difference.

7 Hofstede's cultural dimensions theory

Most psychological research has been conducted on Western samples [44], which has major implications for psychological results because they can only be generalized to Western samples. While some researchers would argue that these psychological processes are universal, there is reason to believe that they depend on socio-cultural contexts and vary between cultures [32, 56]. Therefore, it is important to consider culture when examining psychological processes. Culture is multidimensional and can be difficult to conceptualize. It is an amalgamation of social norms, beliefs, and attitudes consistently influenced by the individuals, groups, and countries the people inhabit [45]. Hofstede [46] created a framework that has been used in multiple fields and is a helpful model to explain how culture might influence an individual's trust in predictive algorithms and their decision to adopt this technology.

The Hofstede model describes cultures in six dimensions. Power distance refers to the extent to which inequality is an issue and if it is addressed. Uncertainty avoidance refers to the level of tolerance a society has for ambiguity and how threatening change is to the culture. Individualism-collectivism refers to how the people within the culture integrate into groups. Masculinity refers to the gender roles within the culture. Finally, long-term orientation refers to the extent to which a culture emphasizes long-term planning and future orientation versus short-term gratification.

Previous research has explored the effect of culture on technology acceptance and innovation. Thatcher et al. [82] found a negative relationship between uncertainty avoidance, power distance, and technology acceptance. In other words, higher levels of uncertainty avoidance and power distance index indicate lower levels of technology acceptance.

One research team conducted a meta-analysis of locus of control using Hofstede's dimensions [14]. They describe agentic-communal goals, which are highly related to individualism-collectivism. Agentic-communal goals refer to how personal accomplishment, success, and power are reinforced in a culture [45]. Perceived control from external circumstances or sources might be distressing for those in individualistic cultures because it threatens their autonomy [14]. With the individualism dimension relating heavily to agentic goals, societies high in individualism might likely have a more internal locus of control. Further, experts might also influence public perceptions.

8 Expert influence

Experts are often perceived as credible authorities because of their extensive knowledge and experience. Research indicates that social movements and public campaigns frequently led or endorsed by experts, can sway public opinion, although the effects tend to be modest and occasionally short-lived [42]. For instance, educational programs on controversial subjects like the death penalty or abortion have been shown to produce small but positive changes in public attitudes [42].

In the realm of emerging technologies such as AI, the influence of experts becomes particularly significant. Public attitudes towards AI are heavily shaped by the views and endorsements of experts. Studies reveal that people turn to experts for insights on the risks and benefits of AI, which can influence their acceptance or rejection of these technologies [9]. Experts help to clarify complex technologies, making them more accessible and less intimidating to the general populace.

Experts play an important role in the public's perceptions of AI. Neri and Cozman [64] found that experts play a crucial role in public risk perception of AI. This indicates that people look to experts to form opinions about new information and contexts. Another study found that trust in government and corporations influences trust in AI [13]. Less is known about the social influence of judges in shaping public trust. Judges are critical stakeholders in the legal system who are likely to influence the adoption and implementation of AI technology significantly. The public's perceptions of judges' trust in AI might influence the public's trust. Although status and authority are elements associated with others trusting new technology or practices, other potential factors could influence public perceptions of trust in AI use in the legal system.

While there is relatively little research on how experts shape public opinions of AI, there is research in other areas documenting the influential power of experts and judges. Regarding COVID, one study found that the public's trust in experts leads to greater uptake of recommended actions during the COVID-19 pandemic [1]. Empirical studies highlight that judges' decisions, particularly in high-profile cases, can significantly influence public opinion. Rulings on civil rights, environmental regulations, and health policies have been shown to shape public attitudes and behaviors over time [93]. Due to their visibility, the Supreme Court's decisions often become focal points for public discourse, reinforcing or altering public norms. As AI is integrated into the justice system, the public will likely rely on judges to form their opinions about AI, underscoring the importance of judicial influence in this context.

9 Current study

There is a dearth of research on the public's perceptions of the use of AI within the justice system. While AI has been shown to be more accurate than human decision-makers, there seems to be a lack of trust in AI decision-makers. We must understand individual differences of the public that might influence their trust in AI technology within the justice system and other domains. We used the following research questions to guide our inquiry into this issue:

10 Research questions

RQ1. Are there differences in the level of trust between the various applications of AI within the criminal justice system?
RQ2. Does perceived judge trust in AI predict trust in the application of AI within the criminal justice system?

RQ3. Does Hofstede's cultural dimensions predict the level of trust in the application of AI within the criminal justice system?
RQ4. Does technology adeptness predict the level of trust in the application of AI within the criminal justice system?
RQ5. Does anthropomorphism predict the level of trust in the application of AI within the criminal justice system?
RQ6. What demographics predict trust in the application of AI within the criminal justice system?
RQ7. What social psychological themes are found within the open-ended question?

11 Methods

11.1 Participants

A total of 150 participants completed an online survey via Prolific, a large crowdsourcing community considered a reliable source for survey sampling in social science research [66]. To establish the most suitable sample size for this particular study, an a priori power analysis was utilized via G*Power (Faul et al. 30). The analysis parameters for within-subject factors were set to detect a small effect size (f = 0.15), with an alpha level of 0.05, a correlation of 0.05, and a power of 0.90.

The sample consisted of 90 females (59.6%), 55 males (36.4%), three non-binary/third gender (2.0%), and two who preferred not to say (1.3%). Participants ranged from 20 to 77 years old (M = 34, SD = 12.1). The majority of participants identified as White (n = 111, 73.5%), followed by Black (n = 17, 11.3%), Asian (n = 8, 5.3%), mixed race (n = 7, 4.7%), Other (n = 5, 3.3%), and American Indian or Alaska Native (n = 2, 1.3%). Many participants had a Bachelor's degree (n = 56, 37.1%), followed by some college but no degree (n = 44, 29.1%), Associate or technical degree (n = 22, 14.6%), Master's degree (n = 12, 7.9%), High school diploma or GED (n = 11, 7.3%), Some high school or less (n = 3, 2.0%), and Professional degree (JD, MD, DDS) (n = 2, 1.3%).

12 Design and procedure

Participants were given instructions that explained that they would be reading about various applications of AI in the legal system. Participants read summaries about the various applications of AI within the legal system. Specifically, they described the use of AI in determining bail (eligibility, fines and fees), sentencing (length, fines and fees), and legal documents (see Appendix A for full descriptions). The bail scenario was the judge using an AI tool to determine bail eligibility and the amount of fines and fees. Sentencing included the judge using an AI tool to determine sentence length and the amount of fines and fees. Finally, the legal document scenario was the judge using AI to help write their legal decision.

After viewing each description, participants were asked about their level of trust within that context. For bail, they were asked about their trust in AI in bail eligibility and bail fines and fees. For sentencing, they were asked about their trust in AI in determining sentence length and fines and fees. They were simply asked about their level of trust in that context for legal documents. Then participants will fill out measures of locus of control, anthropomorphism, Hofstede's cultural dimensions, trust in technology, and technology acceptance.

13 Measures

13.1 Dependent variable

13.1.1 Trust in AI

We determined that trust was the most appropriate dependent measure to assess participants' attitudes toward using AI in various legal contexts. Trust is essential for the acceptance and successful implementation of AI. To measure the level of trust in AI technology in the various legal contexts, we created a single question on a 7-point Likert scale from *not at all* to *completely* "How much would you trust the use of artificial intelligence in ________?".

13.2 Independent variables

13.2.1 Perceived judge trust

We wanted to test how perceived judges' trust influenced participant trust in AI within various legal contexts. To measure the perceived judge's trust we created a single question on a 7-point Likert scale from *not at all* to *completely* "How much would do you believe judge's trust the use of artificial intelligence in its current state?".

13.2.2 Locus of control

To measure *locus of control*, we used the original scale developed by Rotter [70], which details the extent to which individuals believe internal or external forces shape their life. This scale includes 29 statements that measure internal and external locus of control orientation. Participants were asked to select one of two statements that they agreed with more, (e.g., external: "Many times I feel that I have little influence over things that happen to me" or internal: "It is impossible for me to believe that chance or luck play an important role in my life"). We adhere to the scoring procedures described in the scale such that higher scores indicate a more external locus of control (alpha = 0.80; [70]).

13.2.3 Anthropomorphism individual difference scale

To measure anthropomorphism as an individual difference, we used the Individual Differences in Anthropomorphism Questionnaire (IDAQ) scale developed by Waytz et al. [91]. This scale includes 15 items that are rated on a 5-point Likert scale from *not at all* to *very much* (e.g., "to what extent does the average robot have consciousness"). Higher scores indicated higher levels of anthropomorphism (alpha = 0.80).

13.3 Hofstede's cultural dimensions

To measure individual judges' scores on *Hofstede's Cultural Dimensions* a scale made for an individual unit of analysis was used. Yoo et al. [94] developed a psychometrically sound measure of Hofstede's Cultural dimensions to use at the individual level. It includes power-distance, (e.g., power-distance: "People in higher positions should not ask the opinions of people in lower positions too frequently"), uncertainty-avoidance: (e.g., "Instructions for operations are important"), collectivism (e.g., Group welfare is more important than individual rewards"), long-term orientation (e.g., Personal steadiness and stability"), and masculinity (e.g., "It is more important for men to have a professional career than it is for women"). The CVSCALE is a 26-item five-dimension scale that details individual cultural values (alpha = 0.62–0.76).

13.3.1 Technology acceptance

There are three subscales within the Technology Acceptance Model Instrument that each include three items (Teo et al. 81): perceived usefulness (e.g., "Using computers will improve my work," perceived ease of use (e.g., "My interaction with computers is clear and understandable"), and attitudes towards computer use (e.g., "Computers make work more interesting"). These items will be averaged to create a technology acceptance such that higher scores will indicate higher levels of technology acceptance (0.87–0.96).

13.4 Covariates/other variables

13.4.1 Demographics

We asked participants about their demographics, including race, age, education, and gender.

13.5 Use of language enhancement tools

In drafting and revising this paper, we used ChatGPT-4, a large language model trained by OpenAI that has exemplified natural language processing capabilities and generation. ChatGPT was used for assistance in rephrasing and improving the clarity of the writing. It is important to note that while ChatGPT was used to rephrase content, the conceptual development, data collection, analysis, and conclusions drawn in the paper are solely the work of the authors. The use of ChatGPT enhanced the readability of the paper but did not influence academic integrity.

14 Results

We conducted analyses for this study using R, an open-source statistical software. Prior to running the analyses, we analyzed the dataset for missing data. Fortunately, there was only one participant who did not complete the survey. As a result, we removed this participant from the sample, and the remaining data were analyzed, which ensured that we conducted the analyses on a complete dataset, which can help to minimize potential biases and errors.

RQ1. Are there differences in the level of trust between the various applications of AI within the criminal justice system?

A repeated measures ANOVA was performed to compare the effect of Context on Trust. There was a statistically significant difference in Trust between at least two Context groups, $F(3.38, 506.33) = 13.685$, $p < 0.001$ (see Table 1 for post hoc comparisons). In other words, participants had significantly higher trust in bail fees and sentencing fees compared to bail eligibility. They also had significantly lower trust in sentencing compared to bail eligibility. Participants had higher levels of trust for bail fees compared to sentencing. Finally, participants had higher levels of trust in sentencing fees and legal documents compared to sentencing.

RQ2. Does perceived judge trust in AI predict trust in the application of AI within the criminal justice system?

A simple linear regression was used to test if perceived judges' trust significantly predicted trust in the application of AI within the criminal justice system. The fitted regression model was Trust in AI = 1.774 + 0.291*(judge trust). The overall regression was statistically significant ($R^2 = 0.045$, $F(1, 753) = 35.52$, $p < 0.001$). It was found that perceived judge trust significantly predicted trust in AI within the criminal justice system ($\beta = 0.291$, $p < 0.001$). This indicates that the more the participant perceived the judge as trusting the AI, the more they trusted the application of AI within the justice system.

RQ3. Does Hofstede's cultural dimensions predict the level of trust in the application of AI within the criminal justice system?

14.1 Power distance

We fitted a linear mixed model with Trust as the outcome variable, Power Distance as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.07$, $t(148.99) = 0.60$, $p = 0.55$).

14.2 Uncertainty avoidance

We fitted a linear mixed model with Trust as the outcome variable, Uncertainty Avoidance as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.14$, $t(288.14) = 1.38$, $p = 0.17$).

Table 1 Level of trust by AI context in the criminal justice system

Context	M	SD
Bail Eligibility [a]	2.48	0.847
Bail Fees [ab]	2.62	0.843
Sentencing [abcd]	2.24	0.914
Sentencing Fees [ad]	2.66	0.986
Legal documents [c]	2.42	0.962

M = Mean; SD = Standard Deviation. To denote letters marking significant differences, different letters (a, b, c, d) indicate statistical significance. For all variables sharing the same letter, the difference between the means is not statistically significant. If two variables have different letters, they are significantly different.

14.3 Collectivism

We fitted a linear mixed model with Trust as the outcome variable, Collectivism as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.02$, $t(285.34) = 0.20$, $p = 0.85$).

14.4 Long-term

We fitted a linear mixed model with Trust as the outcome variable, long-term as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.10$, $t(283.56) = 0.78$, $p = 0.44$).

14.5 Masculine

We fitted a linear mixed model with Trust as the outcome variable, Power Distance as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.05$, $t(285.20) = 0.62$, $p = 0.53$).

RQ4. Does technology adeptness predict the level of trust in the application of AI within the criminal justice system?

We fitted a linear mixed model with Trust as the outcome variable, tech scale as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = 0.16$, $t(286.17) = 1.31$, $p = 0.19$).

RQ5. Does anthropomorphism predict the level of trust in the application of AI within the criminal justice system?

We fitted a linear mixed model with Trust as the outcome variable, anthropomorphism scale as the predictor, and Subject ID entered as a random effect. Within this model, there were no significant differences ($\beta = -0.01$, $t(284.42) = -0.64$, $p = 0.52$).

RQ6. What demographics predict trust in the application of AI within the criminal justice system?

Due to a lack of power in comparing education groups, they were collapsed into three groups (high school diploma, GED, or less, associate and bachelor, and graduate degrees). There was not a statistically significant interaction between education and context in explaining the trust score, $F(6.73, 498.27) = 0.500$, $p = 0.833$. There was a statistically significant main effect of context ($F(3.37, 498.27) = 6.699$, $p < 0.05$) and education ($F(2, 148) = 3.932$, $p < 0.05$) on the trust score. For education, those who received their high school diploma, GED, or less were more likely to trust AI in various contexts significantly more than those with higher levels of education (Table 2).

RQ7. What social psychological themes are found within the open-ended question?

For the qualitative data analysis, we harnessed the power of ChatGPT-4. Leveraging ChatGPT allowed us to delve into our open-ended responses and extract psychological and sociological themes related to trust in the application of AI within the justice system. ChatGPT provides distinct advantages over other qualitative software. It streamlines qualitative analysis by automating coding and categorization tasks, thereby saving time and enabling researchers to focus on higher-level analysis [16]. Additionally, it mitigates human bias, ensures consistency, fosters iterative and collaborative analysis, and generates fresh insights through interactive dialogues and exploration of multiple perspectives [16].

Table 2 Level of trust by education in the criminal justice system

Context	N	M	SD
High school diploma, GED, or less[ab]	70	3.03	0.93
Associates or Bachelors degree[a]	390	2.45	0.95
Graduate degree[b]	295	2.40	0.93

M = Mean; SD = Standard Deviation. To denote letters marking significant differences, different letters (a, b) indicate statistical significance. For all variables sharing the same letter, the difference between the means is not statistically significant. If two variables have different letters, they are significantly different.

While ChatGPT excels in efficiently analyzing and coding text data, it does have notable limitations. Research indicates that while AI can provide consistency and reduce human error, it can yield varying interpretations based on prompt wording, as evidenced in studies on the trolley dilemma [51]. This variability underscores the necessity of human oversight to validate AI-generated classifications. Furthermore, the use of AI tools like ChatGPT raises significant privacy concerns, as data inputted into these tools may not be securely managed, potentially infringing upon participants' consent agreements [11]. It's important to note that this study did not include any identifiable information within the prompts.

Open-ended responses were entered into the model and set parameters for generating responses related to our research questions. Specifically, ChatGPT was asked it to identify psychological and sociological themes that were stated multiple times within the data. The responses output by ChatGPT were analyzed to identify common psychological and sociological themes within the data. ChatGPT's novelty means it lacks extensive validation and a broad user base, necessitating careful human review. The first author ensured that the quotes received from the data were correct and examined the themes it reported. ChatGPT offered quote examples from the data that are outlined in the text below.

14.6 Social psychological themes

This data set reveals a range of perspectives on the use of AI in the criminal justice system. A thematic analysis of the responses yields several psychological sociological themes, including trust, complexity, nuance, emotion, the need for human involvement, bias and inequality, fairness and equity, the importance of testing and validation, and the complexity of the criminal justice system. Some participants express trust in AI as a means of achieving greater fairness in criminal justice, while others express skepticism or outright opposition. Some participants expressed concerns about bias being amplified, especially against minority groups due to historical systematic racial bias, while others believed it had the potential to make decision-making fairer. The following is a summary of these themes. The following is a summary of these themes.

14.6.1 Trust

Many participants expressed varying degrees of trust and skepticism toward AI in the criminal justice system. Some participants express trust in AI to make fair and unbiased decisions in the criminal justice system (e.g., "I love the idea of AI removing biases from the judicial system"). They believe that AI can be programmed to adhere to a strong directive of equity and that, with proper testing and validation, it is more likely to do a fair and unbiased job than humans (e.g., "if the AI is allowed to use its own interpretation of the laws, I am pretty sure I would trust the AI more than any judge in America"). Others expressed distrust due to the uncertainty of using AI, (e.g., "I think there is so much unknown still about artificial intelligence and how it works that I wouldn't be comfortable with relying solely on it for anything, especially when it has to do with my freedom and livelihood").

14.6.2 Fairness and equity

Many participants expressed interest in the potential of AI to make the criminal justice system fairer and more equitable (e.g., "I love the idea of AI removing biases from the judicial system"). However, some participants expressed concerns that AI may not be able to take into account all relevant factors or may overlook important nuances in individual cases (e.g., "I think there are some cases that can be very alike another, but I also think there are many cases so vastly different from each other that using solely AI to make a logical decision based on patterns could be not as fair").

14.6.3 Complexity and nuance

Several participants expressed the view that AI lacks the ability to account for the human element in criminal justice, such as emotions, extenuating circumstances, and exceptions due to certain circumstances (e.g., "AI can't make this distinction. Everything isn't black and white, there are expectations, and an AI can't make this distinction", "I don't think AI should be used on its own because it can't take into account the human element, like remorse or extenuating circumstances," and "Artificial intelligence cannot account for these emotional outcomes at its current state"). Others expressed concerns about the accuracy and reliability of AI and worried about its use in complex decisions (e.g., "The criminal justice system is very complex, and I would find it difficult to get on board with artificial intelligence making such life-changing decisions. I have little trust in a computer that over-simplifies issues to make a black-and-white decision when there are a plethora of gray shades").

14.6.4 Emotion

Some participants expressed that emotions are a key factor in determining punishment for an individual and that AI cannot assess emotions in decision-making (e.g., "AI can't assess emotions in their decisions. I believe this is a key factor in determining punishment for an individual," and "I think most court cases are influenced by emotions unconsciously, which can result in more positive or more negative outcomes. Artificial intelligence cannot account for these emotional outcomes at its current state").

14.6.5 Need for human involvement

Despite acknowledging the potential benefits of AI, many participants emphasized the importance of human involvement in decision-making (e.g., "I believe that AI on its own will not be feasible at this point in time, but I do think using it to affirm a judge's thoughts or for them to reconsider is helpful"), citing the inability of AI to assess emotions (e.g., "Artificial Intelligence doesn't have empathy"), the value of human compassion (e.g., "I think using it to help inform decisions would be the fairest, but as a criminal, I would prefer for the judge to decide as I think I could appeal to their humanity") and take into account the nuances of individual cases (e.g., "I don't think AI should be used on its own because it can't take into account the human element, like remorse or extenuating circumstances").

14.6.6 Bias and inequality

Many participants expressed concerns about the potential for AI to perpetuate or even amplify existing biases in the criminal justice system (e.g., "My main concern is that the AI would learn the bias that the criminal justice system already has against minorities" and "I can see how it might reduce racial bias sentencing, but it could also make it worse"). They believe that AI has an implicit bias based on its creators and the information it is fed and that this bias could be more harmful to minority groups than having a person with nuance decide (e.g., "AI has an implicit bias based on its creators and the information it is fed, so ultimately it would be more harmful, especially to minority groups than just having a person with nuance decide").

14.6.7 Importance of testing and validation

Many participants emphasized the need for extensive testing and validation of AI systems before they can be implemented in real cases, citing concerns about the potential for AI to introduce new biases or overlook important information in individual cases (e.g., "I think there would need to be extensive testing and validation done before we should trust AI to such a critical piece of the wellbeing of our society").

14.6.8 The complexity of the criminal justice system

Participants acknowledged the complexity of the criminal justice system and expressed doubts about whether AI could fully understand the nuances of individual cases (e.g., 'The criminal justice system is very complex, and I would find it difficult to get on board with artificial intelligence making such life-changing decisions").

15 Discussion

This study explored public perceptions of AI used in bail and sentencing decisions and the writing of legal documents. Participants had varying levels of trust depending on the application of AI within the legal system. Specifically, participants were more likely to trust AI in assessing fines and fees compared to bail eligibility, sentencing length, and legal documentation. This finding suggests that the public might view AI as more objective and less likely to perpetuate bias in financial calculations. This aligns with the personal property relevance phenomenon, such that people are more likely to perceive losses related to personal aspects of their lives as more severe compared to losses to property [69].

In the context of the application of AI within the justice system, this phenomenon could influence people's level of trust in AI. If AI makes decisions concerning an individual's personal well-being (e.g., bail, sentencing), then people are more likely to view it as a threat. However, if AI makes fewer personal decisions (e.g., fines and fees), people are less likely to view it as a threat. Therefore, AI in consequential decisions such as bail and sentencing must be transparent and extensively tested to ensure accuracy and reliability.

One of the main findings of this study is that participants' trust in AI within the legal system is influenced by their perception of judges' trust in AI. Specifically, participants who believed that judges had greater trust in AI were more likely to express their own trust in its use within the legal system. This finding aligns with literature indicating that the public often looks to experts, such as judges, for guidance in uncertain situations [65]. These results highlight the importance of judges' endorsements in shaping public acceptance of AI and underscore the need for ethical guidelines and targeted education for judges. Ensuring judges understand and trust AI technologies can enhance public confidence and acceptance, reinforcing the theoretical framework that public trust is partly derived from the trust expressed by influential experts.

However, it is essential to consider potential counterarguments and alternative explanations for this relationship. For instance, media portrayal of AI and judicial decisions can significantly shape public perception. The media often influences how technologies and judicial actions are viewed, potentially amplifying or distorting judges' endorsements. This broader context underscores the complexity of the issue and suggests areas for further research, such as examining the interplay between media coverage, judicial opinions, and public trust in AI. Addressing these factors strengthens the study's arguments, highlighting the multifaceted nature of public trust in AI and suggesting avenues for future exploration.

It is important to clarify that while judges are authoritative figures within the legal system, their direct role in shaping public opinion towards AI in the legal context is an emerging area of research. The perception that judges endorse AI can lend credibility and legitimacy to its use, aligning with the broader understanding of how expert opinions can influence public attitudes. More research is needed to fully understand the extent and mechanisms of this influence, but our findings suggest that judges do play a significant role in shaping public perceptions of AI in judicial contexts.

By demonstrating trust and using ethical AI, judges may help to increase public trust in AI and facilitate its wider adoption in the legal system. This finding demonstrates the importance of social and psychological factors that influence public perceptions of AI in the legal system. This suggests that gaining support from judges and other legal experts is crucial in the successful development and implementation of AI to minimize adverse effects and maximize positive outcomes of AI.

Interestingly, culture, locus of control, and anthropomorphism were not significant predictors of trust in AI within the legal system. Previous research suggests that culture, locus of control, and anthropomorphism predict trust in AI. One study investigated AI in healthcare and explored how Hofstede's cultural dimensions influence the acceptance of this technology [73]. Of the multiple dimensions studied, the factor that seemed to impact the adoption of AI was uncertainty avoidance, such that high scores were related to non-adoption. One study found that those with a high internal locus of control were more likely to trust their judgment over an AI decision-maker than those who are externally oriented [72]. Studies have found that anthropomorphism is associated with higher trust resilience [23] and greater emotional trust in AI [38]. Further research may be needed to explore these findings in more detail.

Education was a significant predictor of trust in AI within the justice system. Participants who received their high school diploma, GED, or less were more likely to trust AI in various contexts significantly than those with higher levels of education. This indicates that education might play a role in shaping individuals' perceptions of AI. People in higher education are more likely to be skeptical of the potential of AI and aware of its limitations.

Educated individuals might be more skeptical of AI's use within the legal system due to their understanding of its limitations and potential risks. For example, educated individuals might be more aware of its ability to perpetuate biases within the legal system. This increased awareness might increase skepticism toward AI in legal contexts. Another factor that might influence educated individuals' trust in AI is potentially higher expectations for transparency and accountability. Research shows that transparency and accountability are essential for trust [75].

Using ChatGPT, this study also conducted a thematic analysis of the open-ended comments and identified several psychological and sociological themes related to the use of AI within the justice system. These themes included bias and inequality, fairness and equity, the importance of testing and validation, and the complexity of the criminal justice system. Participants expressed concerns about bias being amplified, especially against minority groups due to historical systematic racial bias, while others believed it had the potential to make decision-making fairer. The need for human involvement was also emphasized, suggesting that individuals value the role of human judgment and decision-making in the justice system.

Overall, this study reveals a range of perspectives on using AI in the criminal justice system. The findings suggest that public trust in AI varies depending on the application and that education level may shape individuals' perceptions. The results emphasize judges' significant role in shaping public opinion, particularly regarding the use of AI in the legal system. Specifically, participants who perceived judges as having high trust in AI were likelier to trust AI. This phenomenon aligns with the broader understanding that the public often looks to authoritative figures and experts for guidance in uncertain situations. Given their authoritative status and expertise, judges can lend credibility to new technologies, influencing public acceptance and trust in AI within judicial contexts. This effect underscores the broader influence that expert opinions have on public attitudes, especially in areas where the public may have limited knowledge or experience. The study also highlights the importance of testing and validation and the need for human involvement in decision-making. Further research is needed to explore these findings in more detail and to inform the development and implementation of AI in the justice system.

15.1 Implications

This study can potentially guide future interventions and ongoing education programs for legal professionals. As AI technology becomes more prevalent in the justice system, judges must become familiar with various tools, including decision-making aids and digital evidence. It is essential to provide education and training for legal actors to ensure the successful integration of these tools. To facilitate this process, researchers can investigate how judges develop trust and interact with AI technology, providing valuable insights for legal education and practice. Public trust depends on judge trust; therefore, judges' concerns are relevant and should be considered when implementing these tools.

This research offers insight into public perceptions of using AI within the justice system, which should be considered when implementing this technology. Participants expressed concerns about AI's ability to perpetuate racial bias in bail and sentencing decisions. AI tools used within the courtroom should receive extensive testing to ensure justice. There should also be a process in which defendants can appeal and question the AI process.

If AI is to be used within the justice system, there need to be standardized ethical guidelines. While AI can potentially reduce bias in bail and sentencing, further progress is necessary for developing ethical guidelines. Fjeld et al. [33] identified eight critical themes for AI principles in their review of reports from various continents and organizations. These include privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promoting human values.

15.2 Limitations and future directions

This study has multiple limitations due to the sampling procedures, measures, and design. First, we used a convenience sample of participants using Prolific.ac. Although Prolific is a reliable source for social science research [66], the sample does not represent the national population. Future research could benefit from stratified random sampling, which would allow for the inclusion of underrepresented groups and potentially reveal diverse perceptions of AI in judicial settings. Next, this was an online survey, and participants may have been preoccupied with other tasks, which could have affected their responses. However, respondents were asked to commit to providing thoughtful responses which have been shown to decrease the rate of quality issues and are more effective than other types of attention checks [35]. Future research should consider in-person interviews to allow for more in-depth, nuanced responses.

Third, there were some limitations within the measures. The cultural, locus of control, anthropomorphism, and technology acceptance measures were skewed and not normally distributed, which may explain the nonsignificant results. This is likely due to the sampling method of convenience sampling. The sample was homogeneous and did not collect nuanced differences within the scales. Specifically, for the technology scale, participants who work on a computer were asked to take surveys; therefore, the sample was all technologically savvy. We did run these analyses through log transformations, and this did not make a difference. There is a further limitation with the technology acceptance scale, which traditionally measures how individuals perceive the usefulness and ease of use of technology in their personal lives. However, our study asks participants to evaluate AI's role in judicial decisions, such as sentencing and bail, which may not directly impact their daily lives. This discrepancy could affect their responses, as they are assessing AI's impact on others rather than themselves. Additionally, considering AI's use in the justice system introduces complex issues of fairness and bias, which may influence participants' acceptance of AI differently than in personal or professional contexts.

This distinction highlights the need for further research into how perceptions of bias and fairness specifically affect public opinion on AI in judicial settings.

Another limitation within the measures was the "one-item" trust measure. Future research should obtain a more diverse sample to see a variance within these measures. Further, we used a single-item trust measure. Trust in AI is complicated, and using a single-item measure might not encompass the complex construct. Further, while Hofstede's cultural dimensions were incorporated into the framework, future research could benefit from a more dynamic analysis of how these cultural variables interact with personal attitudes toward technology. An exploration of cross-cultural differences and similarities in perceptions of AI could offer valuable insights into global patterns of trust in technology, enhancing the applicability of the research findings internationally.

Finally, using ChatGPT-4 for qualitative analysis has some limitations. ChatGPT's responses are generated using a predictive algorithm that is trained on a collection of data. The responses it generates might not be accurate, mainly if the user is asking for it to complete a task outside of its trained data. Second, ChatGPT responses lack nuance and complexity. While it can generally create coherent sentences, it may not always understand subtle nuances within the qualitative data. Third, ChatGPT might generate responses that are biased toward a particular perspective. It may not always provide the bigger picture when generating responses. ChatGPT is a transformative resource for qualitative researchers, but researchers should use caution when using ChatGPT for qualitative data analysis and check for quality, accuracy, bias, and relevance.

16 Conclusion

The use of AI in the justice system is becoming increasingly prevalent, but there is a growing concern about ensuring that its decisions are fair and unbiased. AI can reduce bias but also perpetuate biases if designers fail to account for social and cultural factors or train it on biased data. While AI has numerous benefits, such as speeding up decision-making and reducing caseloads, it also has risks, such as the loss of human judgment and empathy and the lack of transparency of AI tools used in court.

There are no standard ethical guidelines for using AI in the justice system. Therefore, it is essential to train judges and decision-makers on how to use AI, including how it works, makes decisions, and interprets its output. Judges must learn how to interpret the output of AI algorithms and identify any biases in the data. This study found that the public's trust in AI is closely related to their trust in judges. Thus, proper training of judges is crucial to maintain public trust in the justice system when using AI.

Researchers must proactively investigate the consequences of AI use and develop ethical guidelines to ensure its fair and appropriate use. To achieve this goal, they should develop mechanisms that promote transparency and accountability in using AI, such as making data sets and algorithms available for scrutiny and enabling appeals or challenges to AI-generated decisions.

The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment covers several principles related to the adoption of AI within the legal system [4]. Principles include maintaining fundamental rights in the development and implementation process, mitigating discrimination, ensuring data privacy and quality, transparency, fairness through explainable methods, and the importance of human intervention. These principles reflect the ethical and responsible use of AI within the legal system.

In conclusion, while AI has the potential to revolutionize the justice system, its use must be approached with caution and transparency to promote procedural justice and maintain public trust. Our study highlights that participants' trust in AI is significantly influenced by their perception of judges' trust in AI, underscoring the need for ethical guidelines and education for judges and decision-makers on AI usage. The responsible integration of AI requires balancing benefits and risks with a clear focus on ensuring fairness, transparency, and accountability. Developing ethical guidelines and educating judges on using AI while maintaining empathy and accountability is crucial. Ultimately, the responsible integration of AI into the justice system demands a careful balance of benefits and risks, with a steadfast commitment to fairness, transparency, and accountability.

Author contributions A.F. wrote the main manuscript text, S.M. worked as an advisor to A.F. and gave extensive feedback and helped with the design of the study. All authors reviewed the manuscript.

Data availability The data for this project is available on OSF: https://osf.io/mzdq3/?view_only=c40d8d3526f749458bdb2338b6f4b032.

Declarations

Competing interests The authors declare no competing interests.

Appendix A

Bail

When someone is charged with a crime, they are presumed innocent until proven guilty. The first step after being charged with a crime is the pre-trial bail hearing. If eligible, one can pay the amount set by the court and remain in the community pending trial. The judge decides if the individual charged with a crime is eligible for bail and for how much. Within bail decisions, judges might use artificial intelligence tools to help determine bail eligibility and amount. These artificial intelligence tools base their decisions on previous historical crime data, which help ensure equity/fairness, as well as predict the likelihood that someone will commit another crime or fail to appear for their court hearing if released.

Sentencing

Once you plead or are found guilty of a crime at trial, you go through the sentencing process. During the sentencing process, a judge determines your punishment typically in the form of some combination of probation, jail/prison time, and/or fines/fees. Within these decisions, judges might use artificial intelligence tools to help ensure equity/fairness and also predict the likelihood that someone will commit another crime when released.

Legal documents

Artificial intelligence has the ability to sort through huge amounts of data and engage in summarizing, interpreting, organizing, and drafting opinions/decisions and other legal documents.

References

1. Ahluwalia SC, Edelen MO, Qureshi N, Etchegaray JM. Trust in experts, not trust in national leadership, leads to greater uptake of recommended actions during the COVID-19 pandemic. Risk Hazards Crisis Public Policy. 2021;12(3):283–302. https://doi.org/10.1002/rhc3.12219.
2. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. In: Ethics of data and analytics. Boca Raton: Auerbach Publications; 2016. p. 254–64.
3. Araujo T. Living up to the chatbot hype: the influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput Hum Behav. 2018;85:183–9. https://doi.org/10.1016/j.chb.2018.03.051.
4. Antinucci M. EU Ethical Charter on the use of artificial intelligence in judicial systems with a part of the law being established on blockchain as a Trojan horse anti-counterfeiting in a global perspective. In: Courier of Kutafin Moscow State Law University (MSAL). 2020; 2: 36–42. https://doi.org/10.17803/2311-5998.2020.66.2.036-042.
5. Barabas, Dinakar K, Ito J, Virza M, Zittrain J. Interventions over predictions: reframing the ethical debate for actuarial risk assessment. arXiv.org. 2018.
6. Bauguess SW. The role of big data, machine learning, and AI in assessing risks: a regulatory perspective. U.S. Securities and Exchange Commission. 2017. https://www.sec.gov/news/speech/bauguess-bigdata-ai.

7. Bauman MJ, Boxer KS, Lin TY, Salomon E, Naveed H, Haynes L, Walsh J, Helsby J, Yoder S, Sullivan R, Schneweis C. Reducing incarceration through prioritized interventions. In: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. pp. 1–8. 2018. https://doi.org/10.1145/3209811.3209869.
8. Brown S. Machine learning, explained. MIT Management Sloan School. 2021; https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.
9. Burstein P. The impact of public opinion on public policy: a review and an agenda. Polit Res Q. 2003;56(1):29–40. https://doi.org/10.2307/3219881.
10. Buskey B, Woods A. Making sense of pretrial risk assessments. National Association of Defense Lawyers. 2018. https://www.nacdl.org/Article/June2018-MakingSenseofPretrialRiskAsses.
11. Canhoto A. Quality and ethical concerns over the use of ChatGPT to analyse interview data in research. Ana Conhoto. 2023. https://anacanhoto.com/2023/04/10/quality-and-ethical-concerns-over-the-use-of-chatgpt-to-analyse-interview-data-in-research/.
12. Chandler J, Schwarz N. Use does not wear ragged the fabric of friendship: thinking of objects as alive makes people less willing to replace them. J Consum Psychol. 2010;20(2):138–45. https://doi.org/10.1016/j.jcps.2009.12.008.
13. Chen YNK, Wen CHR. Impacts of attitudes toward government and corporations on public trust in artificial intelligence. Commun Stud. 2021;72(1):115–31. https://doi.org/10.1080/10510974.2020.1807380.
14. Cheng C, Cheung SF, Chio JHM, Chan MPS. Cultural meaning of perceived control: a meta-analysis of locus of control and psychological symptoms across 18 cultural regions. Psychol Bull. 2013;139(1):152. https://doi.org/10.1037/a0028596.
15. Cherson J. Policy position brief: On pretrial algorithms (risk assessments). The Bail Project. 2022. https://bailproject.org/policy/pretrial-algorithms/.
16. Chesterman P. Leveraging ChatGPT for qualitative analysis: Exploring the power of generative AI. Ethos. 2023. https://ethosapp.com/blog/leveraging-chatgpt-for-qualitative-analysis-exploring-the-power-of-generative-ai/.
17. Chohlas-Wood A. Understanding risk assessment instruments in criminal justice. Brookings. 2020. https://www.brookings.edu/articles/understanding-risk-assessment-instruments-in-criminal-justice/#:~:text=Second%2C%20any%20algorithm%20used%20in,over%20human%20decision%2Dmaking%20processes.
18. Code for America. Los Angeles County DA & Code for America Announce Dismissals of 66,000 Marijuana Convictions, Marking Completion of Five-County Clear My Record Pilot. Code for America. 2020. https://codeforamerica.org/news/los-angeles-county-da-code-for-america-announce-dismissals-of-66-000-marijuana-convictions-marking-completion-of-five-county-clear-my-record-pilot/.
19. Commons ML, Miller PM, Li EY, Gutheil TG. Forensic experts' perceptions of expert bias. Int J Law Psychiatry. 2012;35(5–6):362–71. https://doi.org/10.1016/j.ijlp.2012.09.016.
20. Copeland B. Artificial intelligence. Encyclopedia Britannica. 2022. https://www.britannica.com/technology/artificial-intelligence.
21. Cosmides L. The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. Cognition. 1989;31(3):187–276. https://doi.org/10.1016/0010-0277(89)90023-1.
22. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. Science. 1979;205(4409):997–1003. https://doi.org/10.1126/science.2648573.
23. de Visser EJ, Monfort SS, McKendrick R, Smith MA, McKnight PE, Krueger F, Parasuraman R. Almost human: anthropomorphism increases trust resilience in cognitive agents. J Exp Psychol Appl. 2016;22(3):331. https://doi.org/10.1037/xap0000092.
24. Demuth S, Steffensmeier D. Ethnicity effects on sentence outcomes in large urban courts: comparisons among White, Black, and Hispanic defendants. Soc Sci Q. 2004;85(4):994–1011. https://doi.org/10.1111/j.0038-4941.2004.00255.x.
25. Demuth S, Steffensmeier D. The impact of gender and race-ethnicity in the pretrial release process. Soc Probl. 2004;51(2):222–42. https://doi.org/10.1525/sp.2004.51.2.222.
26. Desai DR, Kroll JA. Trust but verify: a guide to algorithms and the law. Harvard J Law Technol. 2017;31:1–64.
27. Dietvorst BJ, Simmons J, Massey C. Understanding algorithm aversion: forecasters erroneously avoid algorithms after seeing them err. In: Academy of Management Proceedings. Briarcliff Manor, Ny 10510: Academy of Management. 2015; 2014(1): 12227. https://doi.org/10.5465/ambpp.2014.12227abstract.
28. Ehrlinger J, Gilovich T, Ross L. Peering into the bias blind spot: People's assessments of bias in themselves and others. Pers Soc Psychol Bull. 2005;31(5):680–92. https://doi.org/10.1177/0146167204271570.
29. Erebak S, Turgut T. Caregivers' attitudes toward potential robot coworkers in elder care. Cogn Technol Work. 2019;21(2):327–36. https://doi.org/10.1007/s10111-018-0512-0.
30. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 2007;39(2):175–191. https://doi.org/10.3758/BF03193146.
31. Fiechuk A. The use of AI assistants in the courtroom and overcoming privacy concerns. Widener Commonwealth Law Rev. 2019;28(1):135–68.
32. Fiske AP, Kitayama S, Markus HR, Nisbett RE. The cultural matrix of social psychology. In: Gilbert DT, Fiske ST, Lindzey G, editors. The handbook of social psychology. McGraw-Hill; 1998. p. 915–81.
33. Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication, (2020-1). 2020. https://doi.org/10.2139/ssrn.3518482.
34. Gecker J. San Francisco prosecutors turn to AI to reduce racial bias. The Washington Post. 2019. https://www.washingtonpost.com/business/economy/san-francisco-prosecutors-to-use-artificial-intelligence-to-reduce-racial-bias-in-courts/2019/06/12/b37d9a04-8d58-11e9-b08e-cfd89bd36d4e_story.html.
35. Geisen E. Improve data quality by using a commitment request instead of attention checks. Qualtrics. 2022. https://www.qualtrics.com/blog/attention-checks-and-data-quality/.
36. Gilovich T. How we know what isn't so: the fallibility of human reason in everyday life. India: Free Press; 1991.
37. Glaze K, Ho DE, Tsang C. Artificial intelligence for adjudication: the social security administration and AI governance. In: The Oxford Handbook of AI Governance. 2021; Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197579329.013.46.
38. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Acad Manage Ann. 2020;14(2):627–60. https://doi.org/10.5465/annals.2018.0057.

39. Grove WM, Meehl RE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy. Psychol Public Policy Law. 1996;2(2):293–323. https://doi.org/10.1037/1076-8971.2.2.293.
40. Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. Psychol Assess. 2000;12(1):19–30. https://doi.org/10.1037/1040-3590.12.1.19.
41. Hancock PA, Billings DR, Schaefer KE, Chen JY, De Visser EJ, Parasuraman R. A meta-analysis of factors affecting trust in human–robot interaction. Hum Factors. 2011;53(5):517–27. https://doi.org/10.1177/0018720811417254.
42. Harris J. Effective strategies for changing public opinion: a literature review. Sentience Institute. 2021; https://www.sentienceinstitute.org/public-opinion.
43. Harris P. What community supervision officers need to know about actuarial risk assessment and clinical judgment. Federal Prob. 2006;70(2):8–14.
44. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci. 2010;33(2–3):61–83. https://doi.org/10.1017/S0140525X0999152X.
45. Hofstede G. Culture's consequences: comparing values, behaviors, institutions and organizations across nations. USA: Sage Publications; 2001.
46. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. Online Readings Psychol Cult. 2011;2(1):2307–919. https://doi.org/10.9707/2307-0919.1014.
47. Iguazio. What is model accuracy in machine learning? Iguazio. 2023. https://www.iguazio.com/glossary/model-accuracy-in-ml/#:~:text=AI%20accuracy%20is%20the%20percentage,is%20often%20abbreviated%20as%20ACC.
48. Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S. Human decisions and machine predictions. Q J Econ. 2018;133(1):237–93. https://doi.org/10.1093/qje/qjx032.
49. Klingele C. The promises and perils of evidence-based corrections. Notre Dame L Rev. 2015;91:537.
50. Krogue K. Artificial intelligence is here to stay, but consumer trust is a must for AI in business. Forbes. 2017. https://www.forbes.com/sites/kenkrogue/2017/09/11/artificial-intelligence-is-here-to-stay-but-consumer-trust-is-a-must-for-ai-in-business/?sh=6801a857776e.
51. Krügel S, Ostermaier A, Uhl M. ChatGPT's inconsistent moral advice influences users' judgment. Sci Rep. 2023;13(1):4569. https://doi.org/10.1038/s41598-023-31341-0.
52. Kunda Z. The case for motivated reasoning. Psychol Bull. 1990;108:480–98. https://doi.org/10.1037/0033-2909.108.3.480.
53. Lee NT, Lai S. The U.S. can improve its AI governance strategy by addressing online biases. Brookings. 2022. https://www.brookings.edu/blog/techtank/2022/05/17/the-u-s-can-improve-its-ai-governance-strategy-by-addressing-online-biases/.
54. Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process. 2019;151:90–103. https://doi.org/10.1016/j.obhdp.2018.12.005.
55. Malek MdA. Criminal courts' artificial intelligence: the way it reinforces bias and discrimination. AI and Ethics. 2022;2(1):233–45. https://doi.org/10.1007/s43681-022-00137-9.
56. Markus HR, Kitayama S. Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev. 1991;98(2):224. https://doi.org/10.1037/0033-295X.98.2.224.
57. Mayson SG. Dangerous defendants. Yale Law J. 2017;127:490.
58. Misamer M, Signerski-Krieger J, Bartels C, Belz M. Internal locus of control and sense of coherence decrease during the COVID-19 pandemic: a survey of students and professionals in social work. Front Sociol. 2021;6:705809–705809. https://doi.org/10.3389/fsoc.2021.705809.
59. Monahan J, Skeem JL. Risk assessment in criminal sentencing. Annu Rev Clin Psychol. 2016;12:489–513. https://doi.org/10.1146/annurev-clinpsy-021815-092945.
60. Monahan J, Steadman HJ, Silver E, Appelbaum PS, Clark Robbins P, Mulvey EP, Roth LH, Grisso T, Banks S. Rethinking risk assessment. Oxford University Press; 2001.
61. Morantz AD. Mining mining data: bringing empirical analysis to bear on the regulation of safety and health in us mining. West Virgina Law Rev. 2008;111:45.
62. Mossman D. Assessing predictions of violence: being accurate about accuracy. J Consult Clin Psychol. 1994;62(4):783. https://doi.org/10.1037/0022-006X.62.4.783.
63. Neal TMS, Brodsky SL. Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. Psychol Public Policy Law. 2016;22(1):58–76. https://doi.org/10.1037/law0000077.
64. Neri H, Cozman F. The role of experts in the public perception of risk of artificial intelligence. AI Soc. 2020;35:663–73. https://doi.org/10.1007/s00146-019-00924-9.
65. Page BI, Shapiro RY, Dempsey GR. What moves public opinion? Am Polit Sci Rev. 1987;81(1):23–43. https://doi.org/10.2307/1960777.
66. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol. 2017;70:153–63. https://doi.org/10.1016/j.jesp.2017.01.006.
67. Pronin E, Lin DY, Ross L. The bias blind spot: Perceptions of bias in self versus others. Pers Soc Psychol Bull. 2002;28(3):369–81. https://doi.org/10.1177/0146167202286008.
68. Rodrigues R. Legal and human rights issues of AI: gaps, challenges and vulnerabilities. J Respons Technol. 2020;4: 100005. https://doi.org/10.1016/j.jrt.2020.100005.
69. Rossi PH, Simpson JE, Miller JL. Beyond crime seriousness: fitting the punishment to the crime. J Quant Criminol. 1985;1:59–90. https://doi.org/10.1007/BF01065249.
70. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr Gen Appl. 1966;80(1):1. https://doi.org/10.1037/h0092976.
71. Rotter JB, Chance JE, Phares EJ. Applications of a social learning theory of personality. Rinehart and Winston: Holt; 1972.
72. Sharan NN, Romano DM. The effects of personality and locus of control on trust in humans versus artificial intelligence. Heliyon. 2020;6(8): e04572. https://doi.org/10.1016/j.heliyon.2020.e04572.
73. Sharma S, Islam N, Singh G, Dhir A. Why do retail customers adopt artificial intelligence (AI) based autonomous decision-making systems? IEEE Trans Eng Manage. 2022. https://doi.org/10.1109/TEM.2022.3157976.

74. Sherman SJ. Internal-external control and its relationship to attitude change under different social influence techniques. J Pers Soc Psychol. 1973;26(1):23–9. https://doi.org/10.1037/h0034216.
75. Shin D, Zhong B, Biocca FA. Beyond user experience: what constitutes algorithmic experiences? Int J Inf Manage. 2020;52: 102061. https://doi.org/10.1016/j.ijinfomgt.2019.102061.
76. Siau K, Wang W. Building trust in artificial intelligence, machine learning, and robotics. Cutter Bus Technol J. 2018;31(2):47–53.
77. Simmons R. Big data, machine judges, and the legitimacy of the criminal justice system. UC Davis L Rev. 2018;52:1067.
78. Smith V. Maryland, 442 U.S. 735 (1979).
79. Spohn C, Holleran D. The imprisonment penalty paid by young, unemployed black and Hispanic male offenders. Criminology. 2000;38(1):281–306. https://doi.org/10.1111/j.1745-9125.2000.tb00891.x.
80. Starr SB. Evidence-based sentencing and the scientific rationalization of discrimination. Stanford Law Rev. 2014;66:803.
81. Teo T, Milutinović V, Zhou M, Banković D. Technology Acceptance Model Instrument. PsycTESTS. 2017. https://doi.org/10.1037/t64926-000.
82. Thatcher JB, Stepina LP, Srite M, Liu Y. Culture, overload and personal innovativeness with information technology: extending the nomological net. J Comput Inf Syst. 2003;44(1):74–81. https://doi.org/10.1080/08874417.2003.11647554.
83. The US Department of the Treasury. Federal agency data mining report. The US Department of the Treasury. 2009. https://www.treasury.gov/privacy/annual-reports/Documents/FY2008/DataMiningReport.pdf.
84. Turner KB, Johnson JB. A comparison of bail amounts for Hispanics, Whites, and African Americans: a single county analysis. Am J Crim Justice. 2005;30(1):35–53. https://doi.org/10.1007/BF02885880.
85. Twenge JM, Zhang L, Im C. It's beyond my control: a cross-temporal meta-analysis of increasing externality in locus of control, 1960–2002. Personal Soc Psychol Rev. 2004;8(3):308–19. https://doi.org/10.1207/s15327957pspr0803_5.
86. US General Accounting Office. Data mining: Federal efforts cover a wide range of uses. Report to the Ranking Minority Member, Subcommittee on Financial Management, the Budget, and International Security. 2004. https://www.gao.gov/assets/gao-04-548.pdf.
87. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. Philos Technol. 2021;34(4):1607–22. https://doi.org/10.1007/s13347-021-00477-0.
88. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Bus Law Rev. 2019; 494. https://ssrn.com/abstract=3248829.
89. Watson D. The rhetoric and reality of anthropomorphism in artificial intelligence. Mind Mach. 2019;29(3):417–40. https://doi.org/10.1007/s11023-019-09506-6.
90. Western B. Punishment and inequality in America. Russell Sage Foundation; 2006.
91. Waytz A, Cacioppo J, Epley N. Who sees human? The stability and importance of individual differences in anthropomorphism. Perspect Psychol Sci. 2010;5(3):219–32. https://doi.org/10.1177/1745691610369336.
92. Waytz A, Heafner J, Epley N. The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. J Exp Soc Psychol. 2014;52:113–7. https://doi.org/10.1016/j.jesp.2014.01.005.
93. Wihbey J. The Supreme Court, public opinion and decision-making: Research roundup. The Journalist's Resource. 2013. https://journalistsresource.org/politics-and-government/research-roundup-supreme-court-public-opinion/#:~:text=.
94. Yoo B, Donthu N, Lenartowicz T. Measuring Hofstede's five dimensions of cultural values at the individual level: development and validation of CVSCALE. J Int Consum Market. 2011;23(3–4):193–210. https://doi.org/10.1080/08961530.2011.578059.
95. Zhang B, Dafoe A. Artificial Intelligence: American Attitudes and Trends. Center for the Governance of AI Future of Humanity Institute University of Oxford. 2019. https://doi.org/10.2139/ssrn.3312874.