PLOS ONE

RESEARCH ARTICLE

# Judges versus artificial intelligence in juror decision-making in criminal trials: Evidence from two pre-registered experiments

**Eiichiro Watamura**[1]*, **Yichen Liu**[1], **Tomohiro Ioku**[2]

**1** Graduate School of Human Sciences, Osaka University, Suita, Osaka, Japan, **2** Center for International Education and Exchange, Osaka University, Suita, Osaka, Japan

* watamura@hus.osaka-u.ac.jp





**Citation:** Watamura E, Liu Y, Ioku T (2025) Judges versus artificial intelligence in juror decision-making in criminal trials: Evidence from two pre-registered experiments. PLoS ONE 20(1): e0318486. https://doi.org/10.1371/journal.pone.0318486

**Editor:** Bartosz Wojciech Wojciechowski, Uniwersytet Jagiellonski w Krakowie, POLAND

**Received:** October 4, 2024

**Accepted:** January 16, 2025

**Published:** January 30, 2025



**Data Availability Statement:** Our study is registered in the Open Science Framework, from which all materials, including raw data, can be downloaded (https://doi.org/10.17605/OSF.IO/N6TM2).

**Funding:** we confirm that the study was supported by the Japan Society for the Promotion of Science KAKENHI (grant number: 22K03022). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

### Background

Artificial intelligence (AI) is anticipated to play a significant role in criminal trials involving citizen jurors. Prior studies have suggested that AI is not widely preferred in ethical decision-making contexts, but little research has compared jurors' reliance on judgments by human judges versus AI in such settings.

### Objectives

This study examined whether jurors are more likely to defer to judgments by human judges or AI, especially in cases involving mitigating circumstances in which human-like reasoning may be valued.

### Methods

Two pre-registered online experiments were conducted with Japanese participants (Experiment 1: N = 1,735, Mage = 48.4; Experiment 2: N = 1,731, Mage = 48.5). Participants reviewed two murder trial vignettes and made sentencing decisions (1 = suspended sentence; 8 = prison sentence) under two conditions: trials with and without mitigating circumstances.

### Results and conclusion

Across both experiments, participants showed no preference for deferring to human judges' or AI judgments when making sentencing decisions. While suspended sentences were more common in cases with mitigating circumstances, this tendency was unrelated to the judgment source. These findings suggest that jurors do not inherently avoid algorithmic judgments and may consider AI opinions on par with those of human judges in certain contexts. However, whether this leads to improved decision-making quality remains an open question, as objectivity (a strength of AI) and emotional considerations (a safeguard for fairness) may interact in complex ways during juror deliberations. Future research should

**Competing interests:** The authors have declared that no competing interests exist.

further explore how these factors influence juror attitudes and decisions in diverse trial scenarios, taking into account potential biases in existing literature.

## Introduction

The use of artificial intelligence (AI) in courts is accelerating [1, 2]. AI can refine information extracted from testimony and text [3–5], analyze surveillance camera images to identify perpetrators [6], classify investigative materials, and prepare trial transcripts efficiently [7, 8]. It has also proven useful as an assistant to judges, for example, by determining which evidence and testimony are conclusive and reliable to prove facts [9], identifying similar cases, and suggesting sentences based on precedents [10]. Expectations are growing for the realization of robot judges (also known as "AI judges" and "algorithmic judges"), which can replace human judges and make decisions automatically based on vast amounts of case data [11–14].

These benefits are substantial; however, they must be balanced against the unique challenges posed by AI in judicial contexts. Using AI in the judicial domain presents challenges that must be resolved. As AI learns from training data, it may incorporate biases contained in the data [15]. Any discrimination based on race, gender, social background, or other demographic factors created by bias in training data would threaten the fairness of judgments. The supposed "black-box problem" is also a major concern [16, 17]. Although accountability is an essential element for decision-making in court [e.g., 16–20], the process of how the AI arrives at a particular conclusion or judgment is opaque because of the lack of access to the internal workings of the algorithm, the decision criteria, and the learning process [21, 22]. This opacity makes it difficult for AI to meet the accountability criterion at present, although some argue that the opacity issue is not important, as the human mind is similar (e.g., [23]).

Nevertheless, AI offers significant potential to address human limitations, particularly in eliminating cognitive biases and emotional influences that often affect human judgment. These advantages suggest that its use in courtrooms is not only inevitable but also essential for achieving fairer and more efficient trials. People are likely to make judgments based on accessible memories, such as recent or memorable cases (availability heuristic; [24]), or make quantitative judgments according to numerical information given in advance, such as the prosecutor's plea or the amount of damage claimed (anchoring; [25]). Moreover, people often do not consistently judge the same case (noise; [26]). Emotions, such as anger and sadness, can influence judgments, which can change decisions [27–29]. Of course, issues such as dealing with bias in AI training data and ensuring transparency in the decision-making process must be overcome, but by mitigating cognitive bias and emotional influence, AI has the potential to greatly improve the fairness and consistency of judicial decisions. Furthermore, organizing large volumes of documents and evidence using AI would significantly shorten the time required to reach a judgment and reduce litigation delays [30, 31]. AI can also reduce the labor costs and time required to run a courtroom, as automation, especially of simple and repetitive tasks, will make courts more cost-efficient [32, 33]. In addition, AI-powered online platforms and chatbots will make it easier for the public to obtain legal advice and assistance [34, 35], improving access to legal services [36]. Owing to the many potential benefits and those described above, the possibility of AI being introduced into the courts is now realistic [37]. Thus, our focus should not be on whether AI should be introduced into the courtroom but rather on the emerging question of how to use it in courtroom settings successfully.

AI could be used in criminal trials with citizen participation, such as jury and adversarial trials [38]. In these courts, there could be a procedure whereby juries make their decisions

based on the judge's and AI's verdicts. However, it is largely unclear whether jurors are more likely to defer to a judge or an AI, particularly if their decisions are inconsistent. If jurors place too much trust in the AI's judgment or ignore helpful information from the AI and blindly follow the judge's opinion, they could come to a biased decision. The jury's job is to determine the facts and render a fair verdict. If jurors rely too much on the AI or the judge, there is a risk that they will become mere bystanders and lose their ability to make independent judgments. Consequently, for maintaining the fairness and credibility of future justice, it is essential that jurors understand the differences between the opinions of AI and judges and that the degree to which they value one over the other is examined. In particular, we must be prepared for extreme situations in which jurors give more weight to either a judge or an AI when the two are in conflict. Identifying the conditions under which a judge or AI is more likely to be deferred to can clarify the division of roles between the two in a human–machine hybrid system [39] and facilitate the optimization of the system so that one side's opinion is not neglected. This study is a forerunner to this approach.

## Literature review

Do jurors find the judgments of human judges or AI more helpful? Some studies show that one is more trusted than the other, whereas others conclude that there is no difference. These conflicting outcomes have led to a complex and confusing debate [40] regarding trust in AI and human judgment. An experiment conducted with 958 Dutch people found that an automated AI decision-making process was rated as more useful than human experts in important judicial decisions, such as proceedings to initiate a lawsuit [41]. In situations where objective fairness is important, AI is more likely to be considered superior to human experts [42–44]. Despite this high regard for the fairness of AI, several studies have also demonstrated an algorithm avoidance tendency, indicating that people do not consider its decisions as acceptable as those of human experts [45, 46].

However, existing research has primarily focused on civil cases or scenarios in which AI and human judgments do not directly conflict, leaving a gap in understanding how jurors navigate situations wherein AI and human judges provide inconsistent recommendations. In Chen et al.'s [37] experiment, participants rated a human judge's decision as being procedurally fairer than the decision made by a robot judge across three trials (consumer retail arbitration, bail decision, and sentencing decision). Meanwhile, Hayashi et al.'s [47] experiment examined the sentencing decisions of citizens acting as jurors in a trial against a defendant in a robbery-homicide case. The results showed that participants deferred to both expert and AI sentencing requests but deferred more to the opinions of human experts when they wanted a heavier sentence in more crucial decisions. In an experiment by Yalcin et al. [48], participants read a fictitious vignette about going to a local court to divorce their partner and were asked about their trust in an algorithmic judge or human judge and their intention to file a lawsuit. The results showed that participants placed more trust in the human judge and expressed their intention to file the case in a court where a human judge would decide, rather than one where an algorithmic judge would rule.

The reason for avoiding AI's judgments is believed to be that it lacks the human-like ability to consider the subjective factors behind the evidence and law: that is, emotional, moral, and social factors [49, 50]. Indeed, in Yalcin et al.'s [48] divorce litigation experiment, AI was less trusted than in other cases when the partner who was about to leave was experiencing mental health problems. Conversely, there is evidence that if an AI has human-like capabilities, the tendency toward algorithm avoidance disappears. For instance, in an experiment by Watamura et al. [17] using trial clips, a robot judge who was empathetic toward the defendant was

trusted by the participants as much as a human judge. The robot's sentencing decision was also accepted to the same extent as the judge's. These results suggest that AI is avoided because its decision-making is mostly not human-like.

Despite these insights, there is a lack of research examining juror behavior in criminal trials in which AI and human judges present conflicting judgments, particularly in the presence of mitigating circumstances that require nuanced, human-like understanding. To summarize the results of previous studies, AI decisions are likely to be used as frequently as those of judges when human feelings do not need to be considered. However, when such sentiments are expected to be taken into account, a tendency to avoid algorithms is likely to emerge. In the context of procedural justice, studies have shown that the decisions of authority figures believed to understand the feelings of the parties involved are more likely to be accepted [51, 52]. Procedural justice refers to the perceived fairness of the processes that lead to outcomes, emphasizing transparency, impartiality, and the ability to voice concerns [53]. As AI is regarded as lacking the ability to understand emotions to the extent humans do [49, 50], it is unlikely to be trusted in court cases where this ability is required. In a criminal trial with citizen participation, where there are mitigating circumstances requiring human-like competence, the jury will defer to the judge's decision rather than a decision made by the AI. In trials without mitigating circumstances, jurors are likely to defer equally to the judge's and the AI's decisions.

## The present study

This study conducted two pre-registered online experiments to examine whether jurors were more likely to defer to the human judge or AI's judgment when making decisions in criminal trials. In the context of Japan's lay judge system, professional judges and lay judges collaborate to deliberate and decide both the verdict and sentencing. This study assumes such a system, wherein jurors are presented with both a judge's and an AI's opinions to aid their decision-making. This differs from Anglo–American jury trials, which generally separate jurors' verdicts from judges' sentencing responsibilities. Mock jury participants read case vignettes and decided whether the defendant should be given prison or a suspended sentence. They were presented with decisions recommended by a judge and AI to examine which decision participants chose when these decisions conflicted, one recommending a prison sentence and the other a suspended sentence. The outcome measure was the participants' sentencing decision, ranging from 1 (suspended sentence) to 8 (prison sentence). If more weight were given to the judge's decision, the participants would choose a prison sentence under the conditions in which the judge imposed a prison sentence and a suspended sentence under the conditions in which the judge imposed a suspended sentence. No such difference would be seen if the AI were mentioned as much as the judge. Thus, the following hypothesis was tested:

Hypothesis: A judge's judgment is more likely to be deferred to than that of AI in cases with mitigating circumstances.

To establish mitigating circumstances, Experiment 1 used a case in which a defendant who was a victim of domestic violence murdered her husband, and Experiment 2 used a case in which a defendant murdered his mother, who had terminal cancer. These vignettes were selected to reflect criminal cases frequently reported in Japanese media, such as domestic violence and familial homicide, which are both socially and legally significant issues in Japan. By using scenarios that participants might find relatable and relevant, we aimed to enhance the ecological validity of the study and ensure that the cases would engage participants' judgments authentically. This selection aligned with the study's goal of investigating how mitigating circumstances and decision-making sources (AI vs. human judges) influence jurors' decisions.

This study's contribution to the literature lies in examining jurors' preferences for either judges or AI when making decisions by directly manipulating an independent variable that is not good for AI: in this case, the presence or absence of mitigating circumstances. Although Yalcin et al.'s [48] experiment had a similar manipulation, it was conducted in a civil case trial. Thus, to our knowledge, this study is the first to take such an experimental approach in a criminal case.

## Experiment 1

Participants were randomly assigned to one of four groups that combined two conditions of mitigating circumstances (present vs. absent) and two conditions of the decision pattern (judge vs. AI recommending a prison or suspended sentence). The decision patterns were designed to always conflict, ensuring that participants had to decide whether the defendant should receive a prison sentence or a suspended sentence. This setup allowed us to directly test participants' deference to the judge or AI under different conditions.

### Methods

This online experiment, as well as Experiment 2, was approved by the Ethical Review Committee of the Department of Behavioral Sciences, Graduate School of Human Sciences, Osaka University (Approval number: HB023-125), and was conducted based on pre-registered procedures and analysis methods. As described in the subsection on procedures below, informed consent was obtained from participants in electronic form. The pre-registration details and raw data are available on the Open Science Framework.

**Participants.** The target sample size (N = 1,200) was based on Yalcin et al. [48], who used a similar 2 × 2 experimental design to detect small to medium effect sizes (f = 0.10) with 80% power at $\alpha = 0.05$. To account for potential exclusions due to attention checks, we oversampled by 500 participants. Participants were randomly selected from a Japanese Internet research firm's panel of individuals aged 18 years or older. Invitations were mailed to 20,884 participants to ensure sufficient response rates, accounting for an estimated 30% exclusion rate from attention checks (see the next subsection). Recruitment for this study began on December 11, 2023, and ended on December 14, 2023. The panel was demographically unbiased, drawn from all over Japan, and largely representative of the Japanese population. As 98.5% of the Japanese public is ethnically Japanese [54], the authors decided not to collect ethnicity data from the survey participants. Participants were given a reward through shopping points redeemable for Amazon gift cards and other products.

**Procedures.** The procedure followed in this study was based on those of previous studies that used vignettes [47, 48]. Of the panel to whom invitations were sent, those participants who read the instructions and provided informed consent by clicking the "I participate" button were directed to the experimental screen. A total of 1,735 participants—slightly more than the target number—completed the questionnaire (869 men and 866 women, $M_{age} = 48.4$, $SD = 14.9$). First, participants were informed about the task: "As a juror in a criminal trial, decide whether the defendant should be sentenced to prison or given a suspended sentence." Next, they read an introduction to the judge and the AI:

> In this jury trial, both the judge and the AI encourage you to make decisions against the defendant. The judge is a very experienced veteran, and the AI has machine-learned a great deal of case law. Detailed information about the defendant's background, criminal record information, and mitigating circumstances are distributed equally to the judge and the AI.

Next, the participants read a case vignette (see S1 Appendix). The trial vignette, which was a condition with mitigating circumstances, described a female defendant accused of murdering her husband; the woman had been regularly subjected to domestic violence by the victim. Out of a desire to escape the pain, she strangled the victim in his sleep using a rope. The trial vignette for the without-mitigating-circumstances condition described a female defendant who was fed up with her husband, who suffered from a chronic illness, and strangled him in his sleep.

For each condition, half of the participants were asked to read the following description: "The judge ruled that the defendant should be sentenced to prison, and the AI ruled that the defendant should be given a suspended sentence." The other half were asked to read this description: "The judge ruled that the defendant should be given a suspended sentence, and the AI ruled that the defendant should be sentenced to prison." Participants answered the following questions in order:

(1) "If you were present at this trial as a juror, would you decide that the defendant should be given a prison sentence or a suspended sentence?" Participants were asked to answer this question on an 8-point scale (1 = suspended sentence to 8 = actual sentence).

(2) As a manipulation check, the participants were asked to respond to three items related to the mitigating circumstances ("I think there are circumstances in favor of the defendant," "It is too much to put all responsibility on the defendant," and "The defendant is not completely at fault") on a 6-point scale ranging from 1 ("I do not agree at all") to 6 ("I strongly agree,").

(3) Finally, as an attention check to ensure that they had read the vignette properly, the participants were asked to answer "Yes" or "No" to three items describing the details of the vignette (e.g., "The female defendant killed her husband using a knife"). These items were not intended to assess participants' memory but rather to verify their engagement with the experimental materials. This approach was taken to maintain the validity of the responses, as online experiments may include participants who answer without fully engaging with the task.

## Results

The analysis was performed as per the pre-registration. HAD version 17.3 [55] was used for analysis. First, 948 participants who incorrectly answered at least one of the three attention check questions were excluded (see S3 Appendix for results of the following analyses with a full sample). A total of 787 participants (424 women, 363 men, $M_{age}$ = 49.91, $SD$ = 14.68) from the four groups were included in the analysis, ranging from 168 to 233 in each group. A post hoc power analysis using G*Power confirmed that with a small effect size (f = 0.10), $\alpha = 0.05$, and a sample size of 787, this study achieved a power of 0.80. This ensures the sample size was sufficient to detect small effects reliably. Analyses were conducted to confirm the validity of the experimental manipulation of the mitigating circumstances. The reliability coefficients for the three items were sufficiently high ($\alpha = .86$); thus, a t-test was conducted using averaged scores. The results showed that the two groups exposed to the with-mitigating-circumstances condition ($M$ = 4.48, $SD$ = 0.95) considered that there were significantly greater mitigating circumstances than the two groups in the without-mitigating-circumstances condition ($M$ = 3.42, $SD$ = 1.08) ($t(785) = -14.14$, $p < .00$, Hodge's g = -1.02, 95% CI = -1.17, -0.87).

As the experimental manipulation of the mitigating circumstances was valid, an analysis of variance was conducted with the participants' judgments measured on an eight-point scale as the dependent variable, and mitigating circumstances and the decision patterns of the judge and AI as the two independent variables. No interaction was found ($F(1,783) = .49$, $p = .49$, partial $\eta 2 = .00$) (Fig 1). There was a significant main effect of the mitigating circumstances,

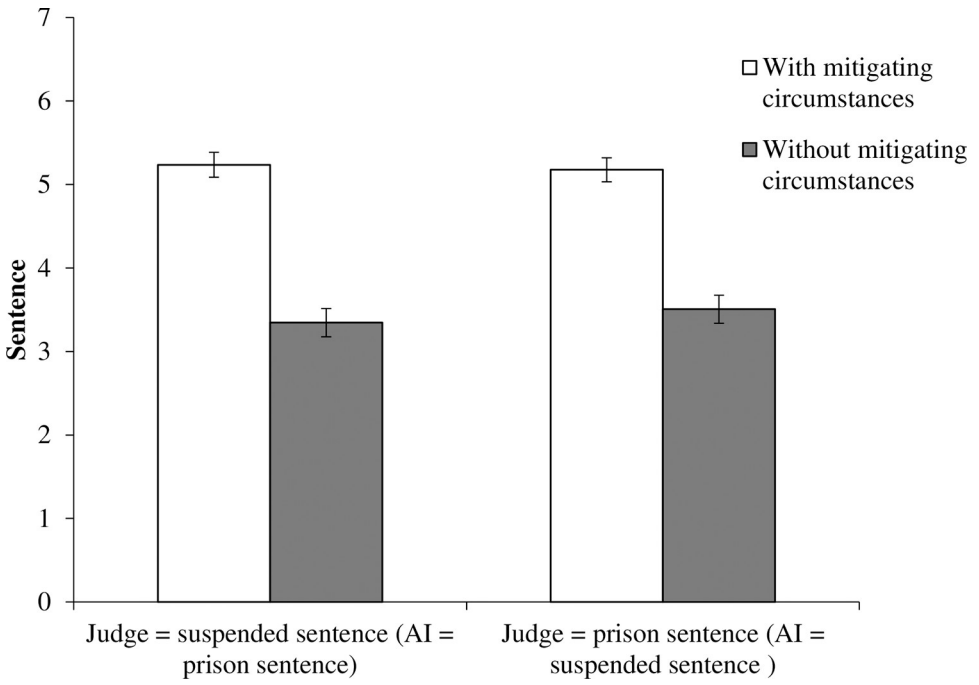


**Fig 1. Jury judgments against defendants in Experiment 1.** Error bars indicate standard errors.

https://doi.org/10.1371/journal.pone.0318486.g001

with prison sentences being more likely to be chosen by the two groups in the without-mitigating-circumstances condition ($M = 5.20$, $SD = 2.09$) than in the two groups in the with-mitigating-circumstances condition ($M = 3.43$, $SD = 2.32$) ($F(1,783) = 127.11$, $p < .00$, Partial $\eta 2 = .14$). No main effect of judge and AI decision patterns was confirmed ($F(1,783) = 0.10$, $p = .75$, partial $\eta 2 = .00$).

## Discussion

Our hypothesis was not supported because the participants' judgments in the with- and without-mitigating-circumstances conditions did not change according to the decision patterns of the judges and AI. We expected the participants to be more likely to defer to the judge than the AI in the with-mitigating circumstances condition. Specifically, we expected that participants' decisions in a case in which the defendant, who had suffered from domestic violence, had murdered her husband would be influenced by the judge's decision and that they would be more likely to opt for a prison sentence if the judge favored a prison sentence and for a suspended sentence if the judge favored a suspended sentence. We predicted that such an advantage would not be found in the absence of mitigating circumstances. The experimental results showed that suspended sentences were more likely to be chosen in the condition with mitigating circumstances than in the condition without mitigating circumstances (i.e., main effect). Notably, there was no main effect of the judges' decision patterns of accepting AI decisions in this study. Previous studies have shown that judges' decisions are more accepted than AI's [37, 48]. This algorithm avoidance tendency [45, 46] might have not been confirmed in the present study due to the particular cases used in the experimental material. For example, participants might have thought that the relationships between the couples were too complex to be understood even by a human (judge). As a result of the judge being less likely to be deferred to, the difference in the ease of deference of the AI might have been reduced, and the algorithm avoidance tendency may not have been confirmed. To examine the generalizability of the findings,

Experiment 2 employed a scenario of murder due to caregiver burnout instead of the maricide scenario used in Experiment 1. Caregiver murder reflects the diversity of morally salient contexts that could influence juror decision-making [56]. Differences in public perceptions of these crimes may offer additional insights into how algorithm avoidance or reliance manifests in varying emotional and moral contexts.

## Experiment 2

In the with-mitigating-circumstances condition of Experiment 2, we used a case in which a male defendant murdered his terminally ill mother. As in Experiment 1, four groups were established, which were exposed to a combination of the mitigating circumstance conditions (with vs. without) and the decision patterns of the judge and AI (the judge sentenced the defendant to prison, and AI gave the defendant a suspended sentence vs. vice versa). Participants were randomly assigned to one of the groups.

### Methods

**Participants.** Our aim was to recruit a sample of 1,700 people. Invitation e-mails were sent to 21,188 randomly selected Japanese panelists aged 18 years or older who were registered with the same Internet research company as in Experiment 1. Recruitment for this study began on December 18, 2023, and ended on December 21, 2023. Participants from Experiment 1 were excluded from this study.

**Procedures.** The 1,731 participants (867 men and 864 women, $M_{age}$ = 48.5, $SD$ = 14.9) who provided informed consent were directed to an experimental screen and instructed to decide whether they would serve the defendant with a prison or a suspended sentence. After reading an introduction stating that the judge and AI were given the same information, they were presented with a trial vignette (see S2 Appendix). The vignette with-mitigating-circumstances condition illustrated a case in which an unemployed male defendant had murdered his elderly mother. The defendant had no choice but to kill his mother because she wanted to die when her terminal cancer worsened. In the trial vignette without mitigating circumstances, a male defendant killed his mother because he was unhappy with being pressured to leave the house. For each condition, half of the participants were told that the judge recommended a prison sentence, and the AI suggested a suspended sentence, whereas the other half were told that the judge wanted a suspended sentence, and the AI proposed a prison sentence. Participants were then asked to rate (1) their judgment against the defendant on an 8-point scale (1 = suspended sentence to 8 = prison sentence), (2) three manipulation check items of mitigating circumstances on a 6-point scale ranging from 1 ("I do not agree at all") to 6 ("I strongly agree,"), and (3) three attention check items (e.g., "The offenders did not have regular jobs at the time of the incident"), to which they answered "Yes" or "No."

### Results

After excluding 871 individuals who answered one or more questions incorrectly during the attention check (see S4 Appendix for results of the following analyses with a full sample), 860 individuals (446 women, 414 men, $M_{age}$ = 50.62, $SD$ = 14.53) remained for analysis (194–238 in each group). A post hoc power analysis using G*Power confirmed that with a small effect size (f = 0.10), $\alpha$ = 0.05, and a sample size of 860, our study achieved a power of 0.83. As the reliability coefficients for the three items on mitigating circumstances were sufficiently high ($\alpha$ = .85), t-tests were conducted using averaged scores. The results showed that the two groups in the with-mitigating-circumstances condition ($M$ = 4.11, $SD$ = 0.94) considered that there were significantly greater mitigating circumstances than did the two groups in the without-

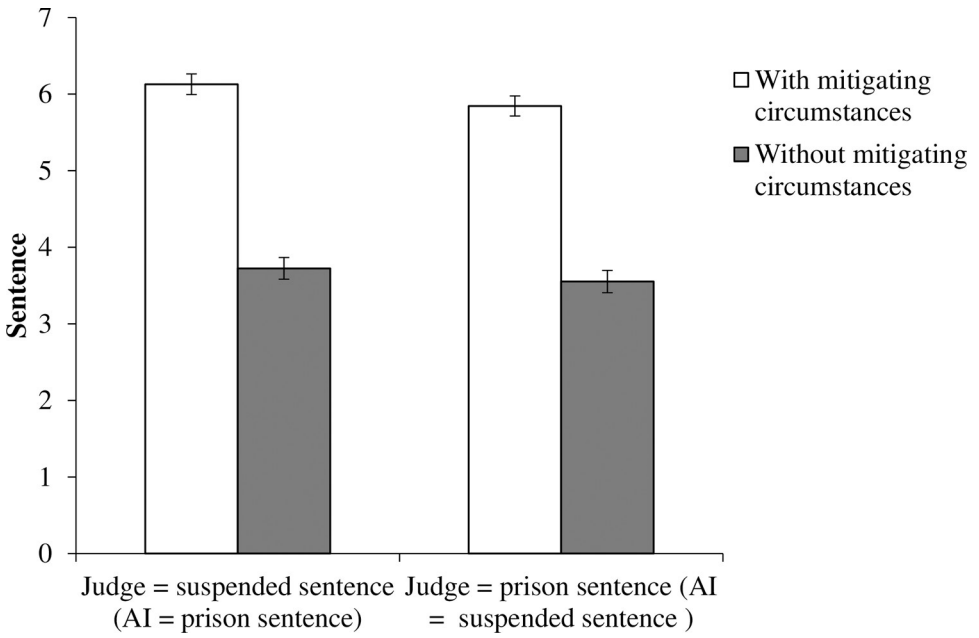


**Fig 2. Jury judgments against defendants in Experiment 2.** Error bars indicate standard errors.

https://doi.org/10.1371/journal.pone.0318486.g002

mitigating-circumstances condition ($M$ = 2.87, $SD$ = 1.09), $t(858)$ = -17.58, $p < .00$, Hodge's g = -1.20, 95% CI = -1.34, -1.06), confirming that we manipulated the mitigating circumstances as intended.

An analysis of variance was conducted with the sentencing decision measured on an 8-point scale as the dependent variable and the two independent variables of the mitigating circumstances and decision patterns of the judge and AI. The results replicated those of Experiment 1. An interaction effect was again not found ($F(1,856)$ = .16, $p$ = .69, partial η2 = .00) (Fig 2). A significant main effect of the circumstances was confirmed in Experiment 2, with the two groups in the without-mitigating-circumstances conditions ($M$ = 5.98, $SD$ = 1.87) being more likely to choose a prison sentence than the two groups in the with-mitigating-circumstances condition ($M$ = 3.64, $SD$ = 2.18) ($F(1,856)$ = 288.79, $p < .00$, partial η2 = .25). A main effect of judge and AI decision patterns was, again, not confirmed ($F(1,856)$ = 2.73, $p$ = .10, partial η2 = .00).

## Discussion

Although we used a different experimental vignette in Experiment 2 to that in Experiment 1, our hypothesis was not still supported. We found no interaction between the presence of mitigating circumstances and the judge's or AI's decision patterns; therefore, our prediction that judges' decisions would more likely be deferred to in the condition with mitigating circumstances was again not confirmed by the results of Experiment 2. Importantly, unlike previous studies [37, 48], Experiment 2 showed no evidence that judges were more likely to be deferred to than AI: Regardless of whether the judge or the AI recommended a prison sentence, the participants' judgments did not change, and their judgments were only influenced by the presence or absence of mitigating circumstances. The lack of interaction between AI and judge decisions, which was also confirmed in Experiment 2 as well as Experiment 1, may be because the participants were equally distant from the factors inherent in both judges. This suggests that the mitigating circumstances themselves played a more central role in influencing participants'

judgments than the perceived humanistic competence or objectivity of the decision-makers. In other words, our results emphasized the possibility that participants prioritized the moral weight of the mitigating circumstances over the decision-making source, whether human or AI.

## General discussion

Are jurors more likely to defer to the judgments of human judges or those of AI? In the future, AI will likely be used in citizen-participatory criminal trials, such as jury and adversarial trials [38, 57]. The future fairness and credibility of justice depends on jurors understanding the differences between the opinions of AI and those of judges and the degree of weight they give to each. Additionally, cultural influences might have shaped jurors' decision-making. In societies in which judicial systems emphasize impartiality and procedural fairness, jurors may place equal trust in AI and human judges, reducing the impact of algorithm avoidance tendencies observed in other studies. By examining the conditions under which each type of judge is likely to be used, we can create a system that can evaluate human judges and AI judges in a human–machine hybrid system without bias [39].

We examined participants' reactions in situations in which the judgments of a human judge and an AI's judgments conflicted. We expected that participants would be influenced by the decision of the one they preferred to defer to more. The presence or absence of mitigating circumstances was the key factor and independent variable. As AI is deemed to lack the human-like ability to consider emotional, moral, and social factors [49], a judge's judgment should be more likely to be deferred to than that of an AI in a trial with mitigating circumstances. Thus, we predicted that jurors would be more likely to choose a prison or suspended sentence if the human judge supported it. However, the results of the two experiments showed that mitigating circumstances were not related to the likelihood of participants deferring to either the judge's decision or that of the AI. Instead, participants focused exclusively on the presence or absence of mitigating circumstances (i.e., main effect): A female victim of domestic violence who murdered her husband (Experiment 1) and a male defendant who murdered his mother, who had terminal cancer and wanted to die (Experiment 2), were more likely to be granted suspended sentences than prison sentences. Thus, although the presence or absence of mitigating circumstances influenced participants' judgments, we found no evidence to predict that the presence or absence of mitigating circumstances made a difference in the likelihood that a human judge or an AI was used. In summary, jurors' tendency to favor more lenient sentences in the face of mitigating circumstances remained unaffected, regardless of who suggested which sentences. These findings contribute to the literature in two ways. First, they showed that jurors did not favor the judge or the AI in making decisions in criminal cases. Second, they identified juror performance in situations where the judge and AI were in conflict. The lack of support for our hypothesis may reflect fundamental differences in how AI and human judges approach complex judicial scenarios. While human judges may rely on moral reasoning and emotional intuition, AI's decision-making is grounded in algorithmic logic, which might not align with jurors' expectations in cases involving nuanced emotional or social contexts.

### Theoretical implications

Notably, this study did not confirm algorithm avoidance tendencies [45, 46], even though the literature has repeatedly shown that AI is not particularly preferred in ethical decision-making situations [58, 59]. Indeed, in Yalcin et al.'s [48] experiment, human judges were preferred in divorce cases with emotional complexity (compared with other cases). However, neither of our experiments confirmed an algorithm avoidance tendency. Unlike Yalcin et al. [48], who used civil case vignettes, this study used criminal case vignettes. In criminal trials, other factors

may offset the effects of algorithmic avoidance. For example, objectivity is likely to be highly valued in criminal trials [60, 61]. Therefore, even if tendencies to avoid algorithms negatively affected the participants' responses in the study, the positive impact of expectations of objectivity could have been offset, which is one of the strengths of AI. As a result, the likelihood of deferring to the AI might have emerged at the same level as the likelihood of deferring to the human judge. Nevertheless, the fact that some studies have demonstrated algorithm avoidance in criminal trials [37, 47] suggests that this explanation may be insufficient. Rather, the results may be due to the unique setting of our study, in which the judgments of the human judge and those of the AI conflicted. In Yalcin et al.'s [48] experiment, participants were assigned to either the AI or judge condition (i.e., between-participant design), which might have confirmed the algorithm avoidance tendency. Alternatively, familiarity with recent technologies, such as ChatGPT, might have reduced negative attitudes toward AI. Another possibility is that there is potentially a large amount of negative data, such as those generated in our study, which have not been reported (i.e., publication bias). Studies that show a tendency to avoid algorithms may be more likely to be reported because people are cautious about AI. Therefore, if more negative data are reported, this would further illustrate people's current attitudes towards AI.

## Practical implications

An important implication of this study is the possibility that jurors would not be biased toward one decision or another if a human–machine hybrid system [39] were introduced in jury trials. The participants were aware of the existence of mitigating circumstances. Nevertheless, our findings suggest that both human judges' and AI's judgments are deferred to equally in criminal trials, even in cases in which humanistic competence to understand mitigating circumstances is particularly necessary. This balance highlights the potential for hybrid systems to reduce undue reliance on either human judges' emotional intuitions or AI's algorithmic objectivity, offering a complementary dynamic in juror deliberations [62]. However, AI's limitations in understanding subjective factors, such as empathy and morality, cannot be overlooked. While AI excels in objectivity and consistency, it lacks the capacity to perceive and evaluate emotional and moral nuances, which are often critical in judicial contexts. For instance, jurors may expect judges to interpret mitigating circumstances with empathy and moral reasoning—abilities that AI currently cannot replicate. This underscores a crucial gap in AI's applicability, particularly in cases that demand a nuanced understanding of human emotions and social complexities. Previous studies have shown that mitigating circumstances reduce perceptions of culpability and influence jurors to recommend more lenient sentences [63, 64]. However, these studies generally did not consider the added influence of a judge's recommendation. Theoretical frameworks, such as the dual-process theory of decision-making [65], suggest that jurors might rely more heavily on a judge's recommendation in emotionally charged cases involving mitigating circumstances. This is because judges, as legal experts, may be perceived as better equipped to balance emotional and rational considerations, particularly when mitigating factors complicate the moral evaluation of the case. We confirmed that jurors valued humaneness, even in AI-supported trials. In cases with mitigating circumstances, people reduce their evaluations of a defendant's culpability or justify a lighter penalty [66, 67]. In our two experiments, the manipulation of mitigating circumstances strongly affected the participants' judgments.

## Limitations and future research

This study has several limitations. First, it remains unclear why there were no differences between participants' deference to AI and that to human judges. Jurors might have viewed AI

as being equally capable of considering mitigating circumstances as human judges, or as noted earlier, another positive benefit of deferring to AI might have been expected. People regard the criteria for moral judgment as being different between AI and human judgments [68]; thus, although the participants deferred to them to the same degree, it was not necessarily for the same reason. Second, the participants in the online experiment were inauthentic. We included a three-question attention check, because online respondents tend to answer questions without reading them [69]. Some participants might have engaged superficially, likely motivated by the financial incentive, which contrasts with those who approached the task seriously, providing thoughtful and authentic responses. We decided to exclude participants who failed at least one attention check to maintain the reliability of our dataset. While this approach ensures the inclusion of engaged participants, it might have inadvertently excluded individuals who genuinely understood the manipulations but overlooked specific vignette details. Future research could explore alternative methods for verifying engagement to better balance data quality with participant inclusion. This should be verified in another setting in which participants answer seriously, such as in a mock guessing experiment. Future research could also investigate the general public's understanding of how AI is trained and operates, as well as the impact of these beliefs on trust and decision-making. This would help clarify whether algorithm avoidance is sometimes justified and how perceptions of AI's capabilities influence behavior. Furthermore, the specific emotional and ethical context of the cases presented in our experiments might have overridden the expected influence of the perceived expertise or empathy of judges and AI. Future research should further explore how contextual factors shape jurors' reliance on different decision-making agents. Another possibility is that the increasing integration of AI technologies in daily life has reduced participants' skepticism toward AI, leading them to treat its recommendations as comparable to those of human judges. Future research should examine whether familiarity with AI or shifts in societal attitudes influence its perceived comparability to human judges.

## Conclusion

Are jurors more likely to defer to the judgments of human judges or AI? We have conducted two experiments to prepare for the introduction of AI in criminal trials with citizen participation. In trials with mitigating circumstances, we predicted that participants would base their verdicts on the recommended decision of a human judge. The results of the experiments showed that the presence or absence of mitigating circumstances affected participants' judgments. However, no evidence supported the prediction that the presence or absence of mitigating circumstances would make a difference in the extent to which they deferred to a human judge or to AI. Consequently, unlike previous studies, we found no evidence of algorithm avoidance. In the context of the lack of known algorithm avoidance in jury trials, our study, using mitigating circumstances as a threshold, raised the possibility that we did not bias jurors toward the decisions of the human judge or the AI. The negative data obtained in this study, where there was no difference in the extent of deference to the judge and AI, suggest that juror decision-making may improve if both are given equal consideration in trials following the introduction of AI. Considering the results of this study and the various benefits that AI brings, we could be more positive about the introduction of AI in criminal courts. However, as this study was an online survey, future research should investigate the ease of mentioning judges and AI in more realistic mock jury experiments and report whether avoidance of algorithms is observed. Furthermore, the findings of this study carry broader societal and ethical implications. As the use of AI in criminal justice systems becomes more prevalent, it is critical to address such issues as the transparency, accountability, and fairness of AI algorithms.

Policymakers should consider the potential risks of algorithmic bias and the need for stringent oversight mechanisms to ensure ethical AI integration in legal settings. Additionally, public education campaigns could be designed to improve understanding and trust in AI-assisted decision-making, thereby reducing skepticism and fostering informed participation in hybrid systems. Future research should explore the long-term impacts of AI on public perceptions of justice and develop frameworks to balance technological advancements with societal values.

## Supporting information

**S1 Appendix. Experiment 1 vignette.**
(DOCX)

**S2 Appendix. Experiment 2 vignette.**
(DOCX)

**S3 Appendix. Jury judgments in Experiment 1 (full sample analysis).** Error bars indicate standard errors.
(DOCX)

**S4 Appendix. Jury judgments in Experiment 2 (full sample analysis).** Error bars indicate standard errors.
(DOCX)

## Acknowledgments

We thank Editage for English language editing.

## Author Contributions

**Conceptualization:** Eiichiro Watamura.

**Formal analysis:** Tomohiro Ioku.

**Funding acquisition:** Eiichiro Watamura.

**Methodology:** Eiichiro Watamura, Tomohiro Ioku.

**Supervision:** Eiichiro Watamura.

**Writing – original draft:** Eiichiro Watamura.

**Writing – review & editing:** Yichen Liu.

## References

1. Carneiro D, Novais P, Andrade F, Zeleznikow J, Neves J. Online dispute resolution: An artificial intelligence perspective. Artif Intell Rev. 2014; 41: 211–240. https://doi.org/10.1007/s10462-011-9305-z
2. Malek MA. Criminal courts' artificial intelligence: The way it reinforces bias and discrimination. AI Ethics. 2022; 2: 233–245. https://doi.org/10.1007/s43681-022-00137-9
3. Gless S. AI in the courtroom: A comparative analysis of machine evidence in criminal trials. Georget J Int Law. 2019; 5: 195–254. Available from: https://heinonline.org/HOL/P?h=hein.journals/geojintl51&i=200
4. Sachoulidou A. Going beyond the "common suspects": To be presumed innocent in the era of algorithms, big data and artificial intelligence. Artif Intell Law. 2023. https://doi.org/10.1007/s10506-023-09347-w
5. Yamada H, Teufel S, Tokunaga T. Building a corpus of legal argumentation in Japanese judgement documents: Towards structure-based summarisation. Artif Intell Law. 2019; 27: 141–170. https://doi.org/10.1007/s10506-019-09242-3

6. Garvie C, Bedoya A, Frankle J. Perpetual Line-Up: Unregulated police face recognition in America. Center on Privacy & Technology at Georgetown Law; 2016. Available from: https://www.perpetuallineup.org.
7. Boella G, Caro LD, Humphreys L, Robaldo L, Rossi P, Van der Torre L. Eunomos, a legal document and knowledge management system for the Web to provide relevant, reliable and up-to-date information on the law. Artif Intell Law. 2016; 24: 245–283. https://doi.org/10.1007/s10506-016-9184-3
8. Shi C, Sourdin T, Li B. The smart court—A new pathway to justice in China? Int J Court Adm. 2021; 12: 4. https://doi.org/10.36745/ijca.367
9. Alves CA. AI assistance in the courtroom and immediacy. In: Morão H, Tavares da Silva RT, editors. Fairness in criminal appeal: A critical and interdisciplinary analysis of the ECtHR case-law. Cham: Springer; 2023. pp. 177–194.
10. Chun AHW. An AI framework for the automatic assessment of e-government forms. AI Mag. 2008; 29: 52. https://doi.org/10.1609/aimag.v29i1.2086
11. Casey AJ, Niblett A. Will robot judges change litigation and settlement outcomes? A first look at the algorithmic replication of prior cases. MIT Comput Law Rep. SSRN Journal. 2020. https://doi.org/10.2139/ssrn.3633037
12. Jongbloed AW, Nakad-Westrate HJ, Herik H, Salem AM. The rise of the robotic judge in modern court proceedings. ICIT: The 7th International Conference on Industrial Technology; 2015; Amman. Jordan Publishing; 2015. https://doi.org/10.15849/icit.2015.0009
13. Morison J, Harkens A. Re-engineering justice? Robot judges, computerised courts and (semi) automated legal decision-making. Leg Stud. 2019; 39: 618–635. https://doi.org/10.1017/lst.2019.5
14. Wang N. "Black box justice": Robot judges and AI-based judgment processes in China's court system. In: IEEE International Symposium on Technology and Society (ISTAS); 2020 November 12–15. IEEE Publications; 2020. pp. 58–65. https://doi.org/10.1109/ISTAS50296.2020.9462216
15. Landers RN, Behrend TS. Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. Am Psychol. 2023; 78: 36–49. https://doi.org/10.1037/amp0000972 PMID: 35157476
16. Deeks A. The judicial demand for explainable artificial intelligence. Columbia Law Rev. 2019; 119: 1829–1850. Available from: https://www.jstor.org/stable/26810851
17. Watamura E, Ioku T, Mukai T, Yamamoto M. Empathetic robot judge, we trust you. Int J Hum Comput Interact. 2023; 40: 5192–5201. https://doi.org/10.1080/10447318.2023.2232982
18. Bell F, Bennett Moses L, Legg M, Silove J, Zalnieriute M. AI decision-making and the courts: A guide for judges, tribunal members and court administrators. Australasian Institute of Judicial administration; 2022.
19. Crawford K, Schultz J. AI systems as state actors. Columbia Law Rev. 2019; 119: 1941–1972. Available from: https://www.jstor.org/stable/26810855
20. Kehl DL, Kessler SA. Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing; 2017. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Available from: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041
21. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data Soc. 2016;3. https://doi.org/10.1177/2053951715622512
22. Von Eschenbach WJ. Transparency and the black box problem: Why we do not trust AI. Philos Technol. 2021; 34: 1607–1622. https://doi.org/10.1007/s13347-021-00477-0
23. Brożek B, Furman M, Jakubiec M, Kucharzyk B. The black box problem revisited. Real and imaginary challenges for automated legal decision making. Artif Intell Law. 2024; 32: 427–440. https://doi.org/10.1007/s10506-023-09356-9
24. Schirrmeister E, Göhring A-L, Warnke P. Psychological biases and heuristics in the context of foresight and scenario processes. Futures Foresight Sci. 2020; 2(2): e31. https://doi.org/10.1002/ffo2.31
25. Bystranowski P, Janik B, Próchnicki M, Skórska P. Anchoring effect in legal decision-making: A meta-analysis. Law Hum Behav. 2021; 45: 1–23. https://doi.org/10.1037/lhb0000438 PMID: 33734746
26. Kahneman D, Sibony O, Sunstein CR. Noise: A flaw in human judgment. London: Hachette UK; 2021.
27. Salerno JM, Phalen HJ. The impact of gruesome photographs on mock jurors' emotional responses and decision making in a civil case. DePaul L Rev. 2019; 69: 633.
28. Oswald ME. How knowledge about the defendant's previous convictions influences judgments of guilt. In: Oswald ME, Bieneck S, Hupfeld-Heinemann J, editors. Social psychology of punishment of crime. Hoboken, NJ: Wiley; 2009. pp. 357–377.
29. Shi J. Artificial intelligence, algorithms and sentencing in Chinese criminal justice: Problems and solutions. Crim Law Forum. 2022; 33: 121–148. https://doi.org/10.1007/s10609-022-09437-5

30. Chen X. Deep learning-based intelligent robot in sentencing. Front Psychol. 2022;13. https://doi.org/10.3389/fpsyg.2022.901796 PMID: 35923731

31. Cui Y. Artificial intelligence and judicial modernization. Cham: Springer; 2020.

32. Barysė D, Sarel R. Algorithms in the court: Does it matter which part of the judicial decision-making is automated? Artif Intell Law. 2024; 32(1): 117–146. https://doi.org/10.1007/s10506-022-09343-6 PMID: 36643574

33. Sourdin T. Judge v robot? Artificial intelligence and judicial decision-making. Univ New S W Law J. 2018; 41: 1114–1133. Available from: https://search.informit.org/doi/10.3316/informit.040979608613368

34. Queudot M, Charton É, Meurs MJ. Improving access to justice with legal chatbots. Stats. 2020; 3: 356–375. https://doi.org/10.3390/stats3030023

35. Ryan F. Delivering legal services without lawyers. In: Jones E, Ryan F, Thanaraj A, Wong T, editors. Digital lawyering: Technology and legal practice in the 21st century. London: Routledge; 2021. pp. 102–135.

36. Morrison A. Artificial intelligence in the courtroom: Increasing or decreasing access to justice? Int J Online Dispute Resol. 2020; 7: 76–93. https://doi.org/10.5553/IJODR/235250022020006001008

37. Chen B, Stremitzer A, Tobia K. Having your day in robot court. Harv J Law Technol. 2023; 36: 127–169.

38. Hayashi Y, Wakabayashi K. Can AI become reliable source to support human decision making in a court scene? In: Companion of the 2017 ACM Conference on Computer-Supported Cooperative Work and Social Computing. Association for Computing Machinery; 2017. pp. 195–198. https://doi.org/10.1145/3022198.3026338

39. Wu T. Will artificial intelligence eat the law? The rise of hybrid social-ordering systems. SSRN Journal. 2019; 119: 2001–2028. https://doi.org/10.2139/ssrn.3492846

40. Lockey S, Gillespie N, Holm D, Someh IA. A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In: Proceedings of the 54th Hawaii International Conference on System Sciences 2021. Institute of Electrical and Electronics Engineers. 2021: 5463–5469. https://doi.org/10.24251/HICSS.2021.664

41. Araujo T, Helberger N, Kruikemeier S, De Vreese CH. In AI we trust? Perceptions about automated decision-making by artificial intelligence. In: AI Soc. 2020; 35: 611–623. https://doi.org/10.1007/s00146-019-00931-w

42. Castelo N, Bos MW, Lehmann DR. Task-dependent algorithm aversion. J Mark Res. 2019; 56: 809–825. https://doi.org/10.1177/0022243719851788

43. 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

44. Mukai T, Yuyama Y, Watamura E. Support for using AI in trials and its determinants. Jpn Soc Inf Knowl. 2023; 33: 3–19. https://doi.org/10.2964/jsik_2023_002

45. Barysė D, Sarel R. Algorithms in the court: Does it matter which part of the judicial decision-making is automated? Artif Intell Law. 2023; 32: 117–146. https://doi.org/10.1007/s10506-022-09343-6 PMID: 36643574

46. Dietvorst BJ, Simmons J, Massey C. Understanding algorithm aversion: Forecasters erroneously avoid algorithms after seeing them err. Acad Manag Proc. 2014;2014. https://doi.org/10.5465/ambpp.2014.12227abstract

47. Hayashi Y, Wakabayashi K, Nishida Y. How sequential suggestions from a robot and human jury influence decision making: A large scale investigation using a court sentencing judgment task. In: Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery; 2023. pp. 338–341. https://doi.org/10.1145/3568294.3580101

48. Yalcin G, Themeli E, Stamhuis E, Philipsen S, Puntoni S. Perceptions of justice by algorithms. Artif Intell Law. 2023; 31: 269–292. https://doi.org/10.1007/s10506-022-09312-z PMID: 37070085

49. Castelo N, Schmitt B, Sarvary M. Human or robot? Consumer responses to radical cognitive enhancement products. J Assoc Consum Res. 2019; 4: 217–230. https://doi.org/10.1086/703462

50. Rai TS, Diermeier D. Corporations are cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel. Organ Behav Hum Decis Process. 2015; 126: 18–26. https://doi.org/10.1016/j.obhdp.2014.10.001

51. Imazai K, Ohbuchi K, Imazai K. Consumer dispute resolution by third party intervention: An experimental study on procedural fairness. Jpn J Soc Psychol. 2003; 19: 144–154. https://doi.org/10.14966/jssp.KJ00003724901

52. Koster N-SN, Van der Leun JP, Kunst MJJ. Crime victims' evaluations of procedural justice and police performance in relation to cooperation: a qualitative study in the Netherlands. Policing and Society. 2020; 30(3): 225–240. https://doi.org/10.1080/10439463.2018.1502290

53. Tyler T. R. What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures. Law & Society Review. 1988; 22(1): 103–135. https://doi.org/10.2307/3053563

54. Minahan J. Ethnic groups of North, East and Central Asia: An encyclopedia (Ethnic groups of the world). Santa Barbara, CA: ABC-CLIO; 2014.

55. Shimizu H. An introduction to the statistical free software HAD: Suggestions to improve teaching, learning and practice data analysis. J Media Inf Commun. 2016; 1: 59. Available from: https://cir.nii.ac.jp/crid/1370286995770875146

56. Itayama A. Comparison between university students and guardians in severe punishment orientation and sentencing judgment. Japanese Journal of Interpersonal and Social Psychology. 2018; 18: 165–171. https://doi.org/10.18910/70554

57. Price WN, Gerke S, Cohen IG. How much can potential jurors tell us about liability for medical artificial intelligence? J Nucl Med. 2021; 62: 15–16. https://doi.org/10.2967/jnumed.120.257196 PMID: 33158905

58. Bigman YE, Gray K. People are averse to machines making moral decisions. Cognition. 2018; 181: 21–34. https://doi.org/10.1016/j.cognition.2018.08.003 PMID: 30107256

59. Sullivan YW, Fosso Wamba S. Moral judgments in the age of artificial intelligence. J Bus Ethics. 2022; 178: 917–943. https://doi.org/10.1007/s10551-022-05053-w

60. Bergman Blix S, Wettergren Å. The emotional interaction of judicial objectivity. Oñati Socio-Legal Series. 2019; 9(5): 726–746. https://doi.org/10.35295/osls.iisl/0000-0000-0000-1031

61. Tyler TR. Procedural justice, legitimacy, and the effective rule of law. Crime Justice. 2003; 30: 283–357. https://doi.org/10.1086/652233

62. Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. Sci Adv. 2018; 4(1). https://doi.org/10.1126/sciadv.aao5580 PMID: 29376122

63. Nadler J, McDonnell M-H. Moral character, motive, and the psychology of blame. Cornell Law Rev. 2011: 97(2): 255–304.

64. Pennington N, Hastie R. The story model for juror decision making. In: Inside the juror: The psychology of juror decision making. 1993. pp. 192–221. https://doi.org/10.1017/cbo9780511752896.010

65. Evans J St BT. Dual-processing accounts of reasoning, judgment, and social cognition. Annu Rev Psychol. 2008; 59: 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629 PMID: 18154502

66. Veiga A, Pina-Sánchez J, Lewis S. Racial and ethnic disparities in sentencing: What do we know, and where should we go? The Howard Journal of Crime and Justice. 2023; 62(2): 167–182. https://doi.org/10.1111/hojo.12496

67. Corwin EP, Cramer RJ, Griffin DA, Brodsky SL. Defendant remorse, need for affect, and juror sentencing decisions. J Am Acad Psychiatry Law. 2012; 40: 41–49. Available from: http://jaapl.org/content/40/1/41.abstract PMID: 22396340

68. Malle BF, Scheutz M, Arnold T, Voiklis J, Cusimano C. Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-robot Interaction. Association for Computing Machinery; 2015. pp. 117–124. https://doi.org/10.1145/2696454.2696458

69. Alvarez RM, Atkeson LR, Levin I, Li Y. Paying attention to inattentive survey respondents. Polit Anal. 2019; 27: 145–162. https://doi.org/10.1017/pan.2018.57