# Legal Affiliates' Views on Algorithmic Decision-Making

**Styliani Kleanthous , Maria Kasinidou**

Cyprus Center for Trustworthy AI, Open University of Cyprus
styliani.kleanthous@ouc.ac.cy, maria.kasinidou@ouc.ac.cy

## Abstract

Procedural fairness in algorithmic decision support systems has been extensively examined across various disciplines. This study investigates how future legal practitioners perceive fairness and accountability in such systems, and how their views diverge from those of computer scientists. In a replication of earlier research, we recruited 150 legal-affiliated crowdworkers via Prolific to: a) rate their agreement with statements related to six fairness constructs across three scenarios, b) define algorithmic fairness, c) identify causes of unfairness, and d) express their views on accountability. Our findings show that legal affiliates' perceptions of algorithmic fairness differ from those in previous studies. Participants often defined fairness as the absence of discrimination, emphasizing the quality of the system's output. They cited 'sensitive attributes' as a primary source of unfairness and held the 'company using or owning the system' accountable when unfairness occurred.

## Introduction

Algorithmic systems play an increasingly prominent role in decision-making (DM) across a wide range of domains, from social media content curation (Rader and Gray 2015; Thorson et al. 2019) to high-stakes applications such as prison release assessments [1] (Angwin et al. 2016) and job recruitment (Mujtaba and Mahapatra 2019). While these technologies promise efficiency and scalability, their deployment has also raised concerns regarding fairness, bias, and accountability. Numerous studies have documented how algorithmic Decision Support Systems (DSS) can perpetuate social inequalities, reinforcing existing societal biases rather than mitigating them (Bozdag 2013). For instance, gender bias has been observed in resume screening algorithms (Chen, Johansson, and Sontag 2018), while racial and gender disparities are evident in AI-generated image search results (Otterbacher, Bates, and Clough 2017; Barlas et al. 2021; Allen 2016; Kyriakou et al. 2020). These disparities highlight the urgent need for critical evaluations of how AI-driven decisions are perceived and whether they align with widely accepted notions of fairness and justice.

Despite these challenges, algorithmic DSS remain essential in high-stakes DM, particularly in legal contexts where fairness, transparency, and accountability are paramount. Legal professionals increasingly interact with AI-driven tools for tasks ranging from legal research and analysis to risk assessments and judicial decision support (Barysė and Sarel 2024). However, concerns persist about overreliance on AI output and gaps in the legal frameworks that govern AI accountability (Chandra and Sanjaya 2023). If AI-driven legal tools are perceived as untrustworthy, they may undermine public confidence in judicial institutions and legal DM processes. Researchers have proposed various models to promote algorithmic fairness (Chouldechova 2017; Dwork et al. 2012) and mitigate biases (Lahoti, Gummadi, and Weikum 2019b,a; Zemel et al. 2013), but their effectiveness depends on stakeholder trust and acceptance.

Understanding algorithmic fairness (Lee 2018; Holstein et al. 2019; Wang, Harper, and Zhu 2020; Woodruff et al. 2018; Binns et al. 2018), transparency (Rader, Cotter, and Cho 2018; Eslami et al. 2019; Tsai and Brusilovsky 2019; Tran et al. 2019), and accountability (Veale, Van Kleek, and Binns 2018; Marcinkowski et al. 2020) between different stakeholder groups is critical. Prior research has primarily focused on developers, policymakers, and the general public interacting with AI-enabled DSS (Grgić-Hlača et al. 2018; Green and Hu 2018; Angerschmid et al. 2022). However, legal professionals, who not only integrate AI tools into legal practice but also play a pivotal role in shaping regulations and ethical guidelines, remain an underexplored group (Kontiainen, Koulu, and Sankari 2022; Said et al. 2023). Given their expertise in interpreting and enforcing legal principles, their perceptions of fairness and accountability are crucial in determining the adoption of such technologies. Importantly, both legal professionals and computer scientists occupy key roles in the lifecycle of algorithmic systems: the former in enforcing laws and shaping legal responses when issues arise, and the latter in designing and deploying the systems themselves. Understanding and comparing the perspectives of both groups is therefore essential for responsible development and governance.

Building on previous work, this study examines the perceptions of algorithmic fairness and accountability among law students and young legal professionals. Our study extends our previous efforts (Kasinidou et al. 2021a; Kleant-



[1] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

hous et al. 2022), which investigated computer science students and recent graduates by specifically exploring how law students and young legal professionals define fairness, identify potential sources of algorithmic unfairness, and assign accountability in cases of biased decisions. To facilitate direct comparisons with our previous findings, we recruited students and recent graduates of law degrees and contrasted their views with prior research (Kasinidou et al. 2021a; Kleanthous et al. 2022). By focusing on this group, we contribute to the growing discourse on AI's role in legal DM, offering insights into how differences in perceptions might shape AI models.

Therefore, we are guiding our analysis by four hypotheses and two research questions, aligning with the work we are replicating (Kasinidou et al. 2021a; Kleanthous et al. 2022):

- **H1:** People who agreed with the decision also believe that the person in the scenario deserved the outcome.
- **H2:** People who found the factors used in the DM process appropriate will also think that the DM process is fair.
- **H3:** People who indicated the DM process was not fair would not trust this system's decision more than a human's decision.
- **H4:** People who agree with the decision would also believe that the factors considered were appropriate.
- **RQ1:** Does the participants' perception of Agreement, Understanding, Appropriateness, Fair Process, Deserved Outcome, and Trust change according to the decision of the system (given the same scenario)?.
- **RQ2:** Are there any differences between Ethnicity Groups and Education Levels in the given constructs and scenarios?

## Related Work

### Procedural Fairness and AI Systems

Procedural fairness, or procedural justice, refers to the perceived fairness of DM processes used to resolve disputes or allocate resources. It is a core component of organizational justice and is closely tied to legitimacy, trust, and compliance with authority (Colquitt and Rodell 2015). Key elements include consistency, impartiality, accuracy, the ability to appeal decisions, and ethical conduct (Leventhal 1980). In the context of AI and automated DM, procedural fairness is crucial for maintaining user confidence and mitigating biases (Binns et al. 2018). Studies have shown that people are more willing to trust AI-driven decisions when they believe the processes behind them are transparent, explainable, and subject to human oversight (Rahwan et al. 2019). For example, in hiring algorithms, applicants may perceive decisions as fairer if they can appeal or receive explanations about the selection criteria (Lee 2018). Similarly, in criminal justice, algorithmic risk assessments must be designed with procedural fairness in mind to prevent systemic biases and ensure just outcomes (Barabas, Dinakar, and Doyle 2019). As AI continues to shape DM in critical areas, integrating procedural fairness principles will be essential for fostering ethical, accountable, and socially acceptable technology.

Extensive discussions in media and academic forums have focused on the challenges of defining algorithmic fairness (Blandin and Kash 2024). Colquitt and Rodell offers a comprehensive review of the constructs of justice and fairness (Colquitt and Rodell 2015), while recent studies show that fairness is closely tied to perceptions of justice (Nagtegaal 2021; Helberger, Araujo, and de Vreese 2020; Ashong Elliot and Arthur 2020). Similarly, trust in the legal system depends on the predictability of judicial decisions, as individuals expect equitable treatment across cases. Thus, perceived trust in procedural justice is critical for the fairness and trustworthiness of judicial decision-makers (Yalcin et al. 2023). The recent integration of DSS in courts affects procedural and perceived fairness, especially given the challenges posed by opaque and unpredictable algorithmic systems. Understanding how legal affiliates, technologists, and the public perceive fairness in this context is crucial, as highlighted by a growing body of research over the past five years.

As AI continues to shape legal practices, discussions on its implications for fairness, accountability, and trust have gained prominence (Greenstein 2022). Traditionally centered on human judgment and procedural fairness, the rule of law faces new challenges as algorithmic DM becomes more entrenched. Scholars have raised concerns that the growing presence of AI in legal systems may diminish fundamental legal principles, particularly when legal professionals and the public perceive algorithmic decisions as opaque or biased (Yalcin et al. 2023). Although AI promises efficiency and cost reduction, trust in human judges remains significantly higher, particularly in those that require emotional sensitivity rather than technical assessments. These concerns highlight the need for a deeper understanding of how legal professionals, who play a pivotal role in shaping and interpreting laws, perceive fairness and accountability.

Existing research suggests that professionals in the legal field exhibit skepticism toward AI-generated content and express concerns about over-reliance on automated DM (Harasta, Novotná, and Savelka 2024). Legal practitioners tend to favor human-authored legal documents over AI-generated ones, despite acknowledging the increasing likelihood of AI-driven legal drafting in the future. Similarly, perceptions of algorithmic fairness vary depending on the stage of judicial DM in which AI is involved (Barysė and Sarel 2024). AI assistance is generally seen as more acceptable in information acquisition rather than in decision selection or implementation, where concerns about fairness become more pronounced. Furthermore, public trust in algorithmic judges is significantly influenced by the nature of the legal case, with lower confidence in AI's ability to handle emotionally complex disputes (Yalcin et al. 2023). Given these findings, our study seeks to expand on this growing body of research by examining how law students (future professionals) conceptualize fairness and accountability in AI-assisted DM.

### Perception of Algorithmic Fairness

Procedural fairness in algorithmic DSS has been of interest within different disciplines (van Berkel, Sarsenbayeva, and Goncalves 2023). Work in the area looked into how people perceive fairness in different scenarios where algorithms are involved (Woodruff et al. 2018; Binns et al. 2018; Lee 2018; Lee and Baykal 2017; Woodruff et al. 2018; Kasinidou et al. 2021a; Kleanthous et al. 2022; Saxena et al. 2019). Although studies are inconclusive, general results reveal that the public perceives algorithmic DM as less fair than human DM (Lee 2018; Marcinkowski et al. 2020; Harrison et al. 2020). In addition, participants in their majority preferred human DM even when they considered the algorithmic model as fair or unbiased (Harrison et al. 2020). They expressed concern about how different factors affect the decision and whether sensitive attributes are taken into consideration (Longoni, Bonezzi, and Morewedge 2019); however, they believe that algorithms might make more objective decisions on some occasions compared to humans (Lee 2018).

Perceptions of algorithmic fairness differ between the general public and legal professionals. Research indicates that the public generally views algorithmic DM as less fair than judges do, although they find algorithmic DM more acceptable during the early stages of information gathering in a case. Conversely, legal professionals consider algorithmic DM to be fairer during the DM phase (Barysė and Sarel 2024). This divergence highlights the distinctive perspectives of legal affiliates compared to the general public regarding the role of AI in DM. Further evidence suggests that people are particularly concerned with whether the decision-maker in court is human or algorithmic. The study found higher levels of trust in human decision-makers (Yalcin et al. 2023). Additionally, Hermstrüwer and Langenbach reported that algorithmic DM was perceived as fairer in scenarios involving significant human involvement, suggesting that the integration of human oversight into algorithmic processes can positively influence perceptions of fairness (Hermstrüwer and Langenbach 2023).

As awareness of algorithmic DM grows, so does skepticism about potential biases in decisions, data, and algorithmic interactions (Araujo et al. 2020; Brown et al. 2019; Malek 2022; Kleanthous et al. 2022). A key challenge for researchers is translating human perceptions of fairness into algorithmic models that consider the "most representative" attributes. The factors and attributes used in these systems can significantly shape perceptions of fairness, both among the public (Grgić-Hlača et al. 2018; Wang 2018; Grgić-Hlača et al. 2018; Wang, Harper, and Zhu 2020) and practitioners (Holstein et al. 2019; Kasinidou et al. 2021a). The importance of specific attributes often depends on the context. For instance, accuracy may be prioritized over equality in some situations, while demographic parity might better align with public notions of fairness in others (Srivastava, Heidari, and Krause 2019). Additionally, factors such as social and altruistic behavior (Lee and Baykal 2017) can influence perceptions, as highlighted by studies showing that attributes considered fair in one scenario may be seen as inappropriate in another (Chen, Johansson, and Sontag 2018; Saxena et al. 2019). This context-dependence underscores the complexity of designing algorithmic systems that align with varying standards of fairness (Green and Hu 2018; Lee 2018; Angerschmid et al. 2022). Studies in the legal domain reveal that practitioners' concerns about algorithmic fairness extend beyond technical issues to include broader implications. These concerns encompass the properties of the technologies themselves, their potential influence on human DM, the absence of robust laws and regulations, gaps in research, and the critical role of the "human factor" in ensuring perceived fairness (Barysė 2022). Addressing these issues is essential to bridge the gap between human perceptions of fairness and algorithmic implementations, particularly in high-stakes contexts such as justice and governance.

### Algorithmic Accountability

As mentioned above, predictability, the sense that cases will be treated equally based on existing procedures and laws, plays a crucial role in people's trust and perception of fairness (Lindquist and Cross 2012). The opacity of algorithmic processes is considered a barrier to predictability and to algorithmic accountability (Diakopoulos 2016; Ramesh et al. 2022; Bogiatzis-Gibbons 2024). Who will be held accountable in case a system behaves unfairly to an individual or a group of individuals in society? A lot of approaches aimed towards providing accountability through transparency (Aysolmaz, Müller, and Meacham 2023; Ramesh et al. 2022; Wang 2018; Nakao et al. 2023). The right to transparency due to regulation attempts has seen several cases in court, demanding public access to information, but still are insufficient to address the harm to human rights that can be potentially caused by the use of algorithms in DM (McGregor, Murray, and Ng 2019).

A study involving the general public found that while perceived transparency had a strong influence on perceptions of fairness and privacy, it was not strongly linked to perceptions of algorithmic accountability (Aysolmaz, Müller, and Meacham 2023). In contrast, when software developers were asked who they believed would be accountable if a system they built produced an unfair outcome, the majority pointed to their teams. They explained that since they were responsible for developing the system, they were also accountable for the outcomes it generated (Kleanthous et al. 2022).

Legal practitioners represent another critical stakeholder group in discussions of algorithmic DM. As DSSs become increasingly integrated into legal practice (Yu and Alì 2019; Barysė 2022), it is essential to understand how legal professionals perceive algorithmic accountability. Their role goes beyond that of end users, they are also actively involved in the regulation, evaluation, and interpretation of algorithmic decisions, including in courtroom settings (Kontiainen, Koulu, and Sankari 2022; Said et al. 2023; Benatti et al. 2024). Gaining insight into their perspectives is crucial for informing both system design and policy development.

## Methodology

To explore how law students and young legal professionals perceive algorithmic fairness and their views on algorithmic accountability, we conducted a crowdsourcing study informed by prior research (Lee 2018; Binns et al. 2018; Kleanthous et al. 2022; Kasinidou et al. 2021a). Specifically, we implemented a replication study to directly compare our findings with our previous work (Kasinidou et al. 2021a) that employed computer science students and young professionals. Participants were presented with three scenarios, adapted from (Kasinidou et al. 2021a), where algorithms made decisions that affected human outcomes. These scenarios were assessed using the same scale employed in the original study (see below), ensuring consistency in measurement and enabling robust comparisons between the findings. This approach provides valuable insights into the unique perspectives of law students and young legal professionals on fairness and accountability in algorithmic DM, shedding light on both similarities and differences across professional domains.

For each scenario, participants were asked to rate their agreement in six statements on a 5-Likert scale, ranging from 'strongly agree' to 'strongly disagree':

1. Agreement(AG): I agree with the decision
2. Understanding(UN): 'I understand the process by which the decision was made'
3. Appropriateness of factors(AP): 'The factors considered in the decision were appropriate'
4. Fair process(FP): 'The decision-making process was fair'
5. Deserved outcome(DO): 'The individual deserved this outcome given their circumstances or behaviour'
6. Trust(TR): 'I would trust this system's decision more than a human's decision

## Scenarios

We employed three different scenarios in this study. One of these scenarios featured three different outcome cases, allowing us to explore how variations in algorithmic decisions influence participants' views.

- **Scenario 1:** Car insurance premiums are dynamically priced based on personal details and driving behavior. This scenario was adopted from (Binns et al. 2018).
- **Scenario 2:** Airline X is using a system for automatically selecting and rerouting passengers on overbooked flights based on the passenger's marital status, number of children the passenger has, whether they are part of a group booking, their age, and gender. Based on this, the system decided to reroute Frank, who was single, traveling alone, and was 55 years old, male, instead of Lisa, who was single, traveling alone, 35 years old, female.
- **Scenario 3:** There are two candidates - Person A and Person B, identical in every way, except their race and loan repayment rates. Both of them have applied for a $50,000 loan to start a business, and the loan officer only has $50,000. This scenario was adapted from Saxena et al.(Saxena et al. 2019) and had three different Cases.
  1. **Case A**: Considering the Gender, Race, and Individual loan repayment rate, the system decided to split the money 50/50 between the two candidates, giving $25,000 to Person A and $25,000 to Person B.
  2. **Case B**: Considering the Gender, Race, and Individual loan repayment rate, the system decided to give Person A $31,818, which is proportional to that person's payback rate of 70%, and give Person B $18,181, which is proportional to that person's payback rate.
  3. **Case C**: Considering the Gender, Race, and Individual Loan Repayment rate, the system decided to give all the money to Person A.

They were also asked to explain using free-text (Q1) *"Was the information provided in the above scenario sufficient?"* and their responses were coded as 'Yes', 'No', 'Unsure'. Next, we asked (Q2) *"How would you define algorithmic fairness?"* (the subtitle read "*Bear in mind that a universally accepted definition of algorithmic fairness does not exist. We just want to know how you personally define algorithmic fairness.*"). Then they were presented with a statement *"In some countries, lawyers (e.g. judges) have access to a system that provides them with a detailed analysis of a case, evaluates arguments and identifies possible outcomes of the case."* and were asked to self-report (Q3) whether they had any previous experience with similar technologies in a 5-point Likert Scale (1, I know absolutely nothing about this - 5, I have tried this or similar technology) and also self-report (Q4) *"How knowledgeable are you regarding Fairness in AI-enabled systems"* in a 5-point Likert scale, (1, Not at all - 5, Very Knowledgeable) and whether they have (Q5) taken *"any training/course on AI"* and (Q6) *"on Fairness, Accountability and Transparency in algorithmic systems"*.

Participants were then presented with the following scenario, which we adapted from (Kasinidou et al. 2021a) to make it context-relevant: *"A company has developed a system to filter and rank CVs, to help hiring managers shortlist the best candidates. The system will rank applicants based on these attributes: Gender, Age, Ethnic Background, Work Experience, Education, Skills, Knowledge, Competencies, Personality Traits."* They were asked to indicate (Q7) *"what would be a possible cause of unfairness in the above system"*. The final questions focused on accountability, and who should be held accountable in case of unfairness. Participants were asked to select one or more of the five statements (Q8): (S1) *The team, which developed the system would be held accountable*; (S2) *The company that owns the system would be held accountable*; (S3) *A company that is using the system would be held accountable*; (S4) *The hiring manager who is using the system would be held accountable*; (S5) *Neither of the above should be held accountable*, and (Q9) *"explain [their] responses"*.

## Participants

The survey was conducted through Prolific[2] filtering those fluent in English and the status of law students and young legal professionals. We recruited 150 participants. The median time for task completion was 19.5 minutes, while the payment was 3.75 GBP per response. The majority were female (68.7%), with 46% in the age group of 18-24, 40.7% between 25-31, 12% between 32-39, and 1.3% above 40 years old. Participants came from diverse backgrounds, with 45.3% identifying as Black, an equal proportion identifying as White, 6.7% as Asian, 2% identified as Mixed ethnicity, and 0.7% identified as Other. In terms of nationality, the majority were from South Africa (66), followed by the United Kingdom (27), Poland (10), Italy (10), Portugal

[2]https://www.prolific.com/

(8), and several other countries, each represented by smaller numbers. Regarding their student status, 16.7% were in their first or second year of undergraduate studies, 32.7% were in their third or fourth year of undergraduate studies, 35.3% were pursuing postgraduate degrees, and 15.3% were recent graduates. In the legal context, 28.7% were working in the legal profession either full or part-time, while the rest were not currently working. Nearly half (46.7%) reported having prior experience with court appearances, including roles such as litigants, observers, defence attorneys, etc. Furthermore, 16.7% were currently appearing in court cases, while 83.3% were not currently actively appearing in court.

### Prior Experience and Knowledge

Interpreting the participants' responses required us first to understand their knowledge (Q4) regarding fairness in AI, whether they have taken any training on AI (Q5) and Fairness, Transparency, Accountability, and Ethics (FATE) in AI (Q6), and whether they had previous experience with DSS in their practice (Q3). Participants' familiarity with Fairness in DSS varied, with 44.7% indicating they were not knowledgeable by selecting the lower end of the scale (1-2, not knowledgeable), 33.3% of our participants opted for the middle (3), indicating an average amount of knowledge, and only 22.0% selected the upper end of the scale (4-5, very knowledgeable). Regarding training on AI, 16.7% reported that they had taken related training, while the majority, 82.7%, had not, and a small part, 0.7%, fell into the "Other" category. Similarly, when asked about training or courses focused on FATE, 9.3% responded affirmatively, with 89.3% reporting that they had not taken such training, and 1.3% answered Other. Finally, participants were asked to report any prior experience or knowledge of decision-support systems (DSS) used in legal contexts—specifically, systems capable of providing detailed case analysis and predicting potential outcomes. The responses revealed that 44% of participants were unaware of such technologies, selecting ratings of 1–2 (indicating no knowledge or familiarity). An additional 19.3% were uncertain, choosing the midpoint (3), while 36.7% indicated some level of experience or exposure to these systems, selecting ratings of 4–5 (suggesting they had tried or were familiar with similar technologies).

### Analysis

To explore participants' perceptions of each construct in Scenarios 1 & 2, and to assess whether these perceptions shift when presented with the same scenario featuring a different algorithmic decision (Scenario 3), we followed the methodology of (Kasinidou et al. 2021a). We aimed to determine whether our results differed from those of (Kasinidou et al. 2021a), which involved CS-affiliated participants.

To test our hypotheses (H1 - H4), we used a series of Wilcoxon signed-rank tests for Scenarios 1 and 2, due to the ordinal nature of the data. To compare the responses received in the three cases of Scenario 3 (RQ1), we followed Friedman's repeated measures non-parametric test, followed by Conover's post-hoc comparison. Mann-Whitney U tests were run to determine differences between groups (Ethnicity and Level of Education) on their perception of the six constructs of justice in the scenarios (RQ2).

To analyze free-text responses in Q1, Q2 & Q7 we conducted a thematic analysis. For Q1 we followed the predefined themes established in (Kasinidou et al. 2021a), and for Q2 and Q7 the themes from (Kleanthous et al. 2022). When deemed necessary to capture emerging themes and nuances within the responses of this study we introduced supplementary themes to have a comprehensive analysis. Two researchers analyzed the responses independently for coding the data based on the themes. They compared and assessed the agreement/consistency between their coding. New emerging themes identified by the researchers were compared and discussed to come to a final consensus. Multiple categories per answer were allowed, and the co-occurrences of themes in Q2 responses were calculated in an attempt to capture the interplay of different themes.

## Quantitative Results

### Perception of Fairness Constructs

To assess data reliability, we conducted a correlation analysis across all fairness constructs, confirming strong correlations in line with (Kasinidou et al. 2021a) (Table 5).

**H1. People who agreed with the decision also believe that the person in the scenario deserved the outcome.** The results show significant differences in the responses of the participants for Scenario 1: $z=2.578$, $p < 0.001$, with participants (30%) selecting options 4 and 5 on the Agreement scale, while 53.3% selected the lower end of the scale (1-2) on the Deserved scale, indicating that they agreed with the decision but the person in the scenario did not deserve the outcome. Different from previous work (Kasinidou et al. 2021a), in Scenario 2 we did not get significant results.

**H2. People who found the factors used in the DM process appropriate will also think that the DM process is fair.** The results show significant differences between the responses of the participants in both scenarios (Scenario 1: $z=5.429$, $p < 0.001$; Scenario 2: $z=3.972$, $p<0.001$). In Scenario 1 participants (50%) selected the upper end of the Appropriateness scale indicating that the factors used were appropriate however, 46% selected the lower end of the Fairness scale indicating that the decision was not fair. Qualitative results indicated that participants extensively discussed how the decision taken was unfair and this was a new theme added. On the other hand, in Scenario 2, participants (61.3%) indicated that the factors considered were not appropriate by selecting the lower end of the scale while indicating (68.7%) that the DM process was not fair. Qualitative results for scenario 2 confirmed this result.

**H3. People who indicated the DM process was not fair would not trust this system's decision more than a human's decision.** Statistically significant results were found for Scenario 1: $z=4.210$, $p<0.001$ but not for Scenario 2. Looking at the participants' responses in Scenario 1, 46% selected the lower end, and 35.33% selected the upper end of the scale for Fairness. 58% selected the lower end of the scale for Trust, while only 20% selected the upper end of the scale for Trust, meaning that the majority would not trust this system's decision more than a human's decision. For Scenario 2, 69% selected the lower end of the scale for Fairness, while 73.33% selected the lower end for Trust.

**H4. People who agree with the decision would also believe that the factors considered were appropriate.** The results show statistically significant differences in the responses of the participants in Scenario 1 (z=-5.191, p<0.001). 50% of the participants believed that the factors considered were appropriate and selected the upper end of the scale on Appropriateness, while only 30% agreed with the system's decision by selecting the upper end of the Agreement Scale. In Scenario 2, 65.33% participants selected the lower end of the Agreement scale, while 61.33% selected the lower end of the Appropriateness scale, indicating in their majority that they do not agree with the decision and that the factors were not appropriate. The qualitative results confirm this finding, mentioning also other factors.

**RQ1. Does the participants' perception of Agreement, Understanding, Appropriateness, Fair Process, Deserved Outcome, and Trust change according to the decision of the system (given the same scenario)?**. Significant differences were found for all constructs (see Table 5) on Friedman's repeated measures test. Conover's post-hoc tests showed that participants perceived the decision in Case A as the most just, while the decision in Case C was the least. In addition, comparing their responses in question Q1 in all three cases in Scenario 3, we observed significant statistical differences ($x^2(2)$=37.587, p <0.001) with the post-hoc test revealing that participants felt that the information provided in Case B was perceived as sufficient. 64.7% indicated sufficient information provided in Case B, 54.7% in Case A, and 33.3% in Case C. Similar to earlier work (Kasinidou et al. 2021a), we did not find any differences in the participants' responses between self-reported gender groups. Previous training and knowledge on topics related to algorithmic DM did not have an impact on their responses.

**RQ2: Are there any differences between Ethnicity Groups and Education Levels in the given constructs and scenarios?**

**Ethnicity Group Differences** Since the literature has been reporting incidences of unfairness involving people of non-white ethnic background, we were interested in investigating whether there are any differences in the perception of the participants involved in this study, according to their self-reported ethnicity. The majority of the participants reported being Black (45.3%) or White (45.3%), thus, we ran a series of Mann-Whitney U tests to determine differences between the two groups. Distributions of the engagement scores between black and white participants were similar as assessed by visual inspection in all cases. For all constructs and scenarios no statistically significant results were observed, except from Scenario 3 Case A (see Table 5). The results indicate that people who self-identified as white were more negative compared to self-identified black participants for all constructs.

**Level of Education Differences** A number of the participants (127) reported their status as a student. In total, 74 undergraduate and 53 postgraduate student responses were analyzed. Using a series of Mann-Whitney U tests we explored the differences between the groups. Distributions of the engagement scores for the two groups were similar, as assessed by visual inspection. For Scenarios 1, Case B, and Case C, we did not find any differences between the two groups. In Scenario 2, there are marginal statistically significant differences for the constructs of Fairness, Deserved Outcome, and Trust (see Table 1). In Scenario 3 Case A, there are differences in the responses between undergraduate and postgraduates in Agreement and Trust (see Table 1).

## Qualitative Results

### Was the information Sufficient?

***Scenario 1.*** In Scenario 1, 106 participants agreed that the information provided was sufficient (Q1), 40 did not agree, and 4 were unsure. 59 elaborated on Q1, referring to one or more of the five adopted themes (no participant referred to Human/Company Policy) and the one newly established theme (Table 2). The most common theme was **Missing Factors**, appearing in 24 responses which referred to *"other factors [that] could have contributed to that outcome"* (participant 10 - p10) such as *"cause of [the] accidents"* (p27), *"frequency of exceeding speed limits"* (p47), *"the car model, the type of place the customer lives"* (p51) and *"client's health like eye tests"* (p64). Surprisingly, some often brought up the importance of taking other factors into account, even though they initially agreed that the information was sufficient. Some asked for **Specific Information** which seemed to be missing from the scenario, such as how much each tier costs (p32, p46) or more information about the accident (p43), as well as more information about how similar was the case of Claire and Sarah (p132, p141). They only inquired about the values of these factors and did not indicate that the system should consider other factors. Others referred to the **Similar Cases** on which the decision was based with some noting that the information was sufficient, assuming that the *"data was compared against [the example] that was quite similar to hers"* (p40). Others questioned *"how the [case used] is relevant"* with Sarah's case (p134) and noted that *"the comparison with Claire was not enough to take any conclusions"* (p95). A few discussed the **Process** followed arguing that *"the system does not explain the reason of the decision"* (p38) or *"reveal the specific factors that led to Sarah's disqualification"* (p117). Eight discussed the newly emerged theme **Unfair** noting that *"the decision to disqualify Sarah was irrational and unfair"* (p30).

***Scenario 2.*** In scenario 2, 73 participants agreed that the information provided was sufficient, 69 disagreed and 8 were unsure, with 77 elaborating on their responses, discussing one or more of the four adopted themes as well as the two newly emerged themes (see Table 3). The most common themes were **Age** and **Gender**. Most disagreed with the use of age (p7, p70) and/or gender (p53, p79), pointing out that *"[they] should not be in consideration for airlines nor is it relevant"* (p150). Even though some questioned *"why an older man is more likely to be rerouted than younger woman"* (p34) others agreed saying that *"it [is] better to reroute an older male customer instead of a young female customer"* (p16). The third most common theme was the

| Constructs | Mann-Whitney U | p | Median Under. | Median Post. |
|---|---|---|---|---|
| Scenario 2 | | | | |
| Fairness | 2360.5 | 0.040 | 2.000 | 2.000 |
| Deserved | 2383 | 0.031 | 2.000 | 1.000 |
| Trust | 2330 | 0.053 | 2.000 | 1.000 |
| Scenario 3 Case A | | | | |
| Agreement | 2390 | 0.027 | 4.000 | 4.000 |
| Trust | 2396 | 0.030 | 3.000 | 3.000 |

Table 1: Differences between participants' level of education

| Theme | Description | # |
|---|---|---|
| Missing Factors | Not considering all the appropriate factors | 24 |
| Specific information | Specific value of a factor missing from the given scenario | 16 |
| Similar cases | Comparison with similar cases, data used to train the model | 15 |
| Process | Procedures followed by the model; features' weights | 9 |
| *Unfair | The decision taken was unfair | 8 |
| Human/Company policy | Deferring to humans, following company's policy | 0 |

Table 2: Themes for Scenario 1. Themes with * are the themes added in this study.

**Factors**. Participants argued that *"other factors more important to be taken into consideration"* (p148), such as health condition (p131), reason(p113), and the urgency (p100) of the travel. Some pointed out *"the factors considered were not appropriate"* (p141). Some also discussed the **Process** followed by the system noting that *"[they] cannot understand the logic behind the decision at all"* (p83) *"[or] why the system prefers younger passengers"* (p2).Others argued that the decision (p106) or the factors used (p117) were **Discriminatory**. A few noted that **Specific Information** was missing from the description of the given scenario.

***Scenario 3.*** In Scenario 3, the analysis grouped the three cases to evaluate how different outcomes influenced participants' perspectives. In Case A, 82 participants agreed that the information was sufficient, 58 indicated insufficient and 10 were unsure. In Case B, 97 expressed satisfaction with the provided information, 46 found it lacking, and 7 were uncertain. In Case C, 50 indicated sufficient information, while 91 disagreed and 9 were uncertain. Case A had 73 responses, Case B had 57, and Case C had 79 responses that elaborated their response. A new theme emerged from the collected responses (refer to Table 4). In Case A, the majority (29 of 73) asked about **Specific Information** which were not included such as *"which race and gender the applicants were"* (p144, Case A [p144/A]) or the *"loan repayment rates"* (p138/A) however, only 11 (of 57) in Case B and 12 (of 79) in Case C mentioned this theme. In Case B, the most common theme was **Race/Gender**, which was also a popular theme in Case A (21 of 73) and Case C (25 of 79). Some raised questions about the use of race and gender,while others argued that these factors were irrelevant to the decision (p54/A, p16/B, p34/C) and should not be used (p100/A, p49/B, p37/C). The **Process** was the most common theme in Case C, only 10 discussed this theme in Cases A and B. Some noted that *"[they] don't know exactly why this is the result"* (p18/C) saying that *"[the system] does not explain in detail the reasons of the decision"* (p38/C). They also referred to **Factors** that should be considered such as *"salaries"* (p59/A), *"the credit history"* (p46/B), *"job, loan repayment ability and what was the reason for the loan"* (p53/A). Others just stated that *"important factors were not included"* (p62/B). A few in Case B and Case C noted that the decision or the system is **Unfair**, arguing that *"everyone should be treated the same"* (p45/B) and that *"the outcome of this scenario was entirely biased"* (p44/C). Overall, 8 responses fell under the **Other** theme, which includes responses that do not fall under any of the other themes.

## Defining Algorithmic Fairness

In defining fairness (Q2), participants referred to the 12 themes adopted from (Kleanthous et al. 2022) (see Table 6a). The majority defined algorithmic fairness in opposition to what is unfair (i.e. **Biases and/or Discrimination**); *"a system that does not discriminate unfairly"* (p30). Others noted that *"[f]airness in algorithms aims to eliminate or minimize the potential for bias"* (p10). One participant also highlighted that *"[t]here will always be some bias in algorithms because of the developers behind the technology who indirectly, or unknowingly insert their own views and biases in the technology"* (p28). Out of the 51 responses discussing Bias/Discrimination (Table 6b): 16 responses also discussed the Outcome/Decision referring to the biased decision; 19 also discussed the Objective Factors concerning the factors that should be considered to eliminate/minimize biases; and 9 also discussed Emotional/Moral/Ethical/Norms.

The second most common theme was **Decision/Outcome**. Participants often referred to algorithmic fairness as *"algorithms [that] make decisions that are just and do not cause harm to individuals or communities"* (p18). Some argued that there should be *"consistent"* (p113), *"reasonable and fair"* (p131) outputs. The most common co-occurrence was with Bias/Discrimination and Objectives Factors. The responses often seemed to imply that the 'correct' factors would lead to the 'best' outcomes, sometimes saying it ex-

| Theme | Description | # |
|---|---|---|
| Age | Consideration of age in the decision | 34 |
| Gender | Consideration of gender in the decision | 32 |
| Factors | Consideration of irrelevant factors and/or missing important factors | 21 |
| Process | Procedures followed by the model; features' weights | 16 |
| *Discrimination | The decision made was discriminatory | 12 |
| *Specific information | Specific value of a factor missing from the given scenario | 6 |
| Other | [falls outside of the established themes] | 1 |

Table 3: Themes for Scenario 2. Themes with * are the themes added to this study.

| Theme | Description | A | B | C |
|---|---|---|---|---|
| Specific information | Specific value of a factor missing from the given scenario | 29 | 11 | 12 |
| Factors | Consideration of irrelevant factors and/or missing important factors | 24 | 11 | 21 |
| Race/Gender | Consideration of race and/or gender in the decision | 21 | 19 | 25 |
| Process | Procedures followed by the model; features' weights | 10 | 10 | 30 |
| *Unfair | The decision made was unfair and/or discriminatory | – | 8 | 18 |
| Other | [falls outside of the established themes] | 2 | 4 | 2 |
| Same as above | Same answer as the previous case(s) | 1 | 1 | 3 |

Table 4: Themes for Scenario 3 (Case A, Case B and Case C). Themes with * are the themes added in this study

plicitly: *"finds an appropriate decision without discriminating any persons by focusing only on factors paramount to the decision"* (p119). The third most common theme was **Objective Factors**.discussing the importance of *"taking into account all appropriate factors"* (p53). Some noted that the algorithm should only use relevant factors (p2, p35, p128) for the goal of the system. Others mentioned **Methods/Rules** highlighting the importance not only of choosing the appropriate factors but of knowing *"how they interrelate"* (p6). 8 of the 13 responses discussing Methods/Rules also discussed Training Data noting that algorithms that *"evaluating a wide database makes fair decisions based on probability and societal conventions"* (p137). 26 responses mentioned the **Emotional/Moral/Ethical/Norms** theme defining algorithmic fairness based on established emotional, moral, ethical, or normative principles. A common sentiment was that a fair algorithm would make *"ethical and equitable treatment of individuals or groups"*(p9), *"treating every individual equally"* (p91), and *"[using] criteria that a big part of society accepts as being morally correct"* (p95).

Some discussed the **Training Data** theme. Often co-occurring with Decision/Outcome (8) and Biases/Discrimination (7), this theme included responses noting that *"the aim [of algorithmic fairness] is to make unbiased decisions based on data provided"* (p51). Nineteen mentioned **Demographics** with the majority mentioning specific factors such as *" gender, race, marital status, religion, belief and culture"* (p23). 8 of the 19 also mentioned Biases/Discrimination, noting that *[the algorithm would be] biased or discriminating against specific groups based on things like race/-gender"* (p18). Responses in **Human Intervention** mentioned the need to include humans in the DM process (p32, p49), while others noted that *" [algorithms can be] influenced by humans"* (p80). Responses in **Context** highlighted the fact that *"fairness is always depending on the viewer's perspective"* (p56). The rest of the themes received only a few responses. The **Explainability/Transparency** noted that it is important to be *"clear why the algorithm made its decision"* (p147). The sentiment of the **Disadvantaged Groups** theme also appeared in the Demographic Characteristics theme. 25 responses fell under the catch-all **Other** category, which includes thoughtful responses that do not mention the other themes.

### Causes of Unfairness

When asked to indicate potential causes of 'unfairness' (Q7), participants referred to one or more of the four adopted themes however, they also discussed two newly established themes (see Table 7). Most referred to the use of **Sensitive Attributes** such as gender, ethnic background, and age, stating that using these can lead *"bias towards particular demographics"* (p90). 9 mentioned the **System/Model** itself as a cause of unfairness, for instance when *"[g]iving too much weight to gender and ethnicity instead of work experience and skills."* (p4). Others noted that the **Lack of appropriate factors** used can be a cause of unfairness. This theme did not exist in the initial set of themes but emerged in this study. Responses mentioned that *"[there] are much more important and work-related [attributes] such as experience, skills or education"* (p36) that the system should take into account. Five discussed the other newly emerged theme **No Human Judge** noting that *"the system is not able to know the character traits of a person who would make an ideal candidate for a position"* (p31). A few believed that **Human Influence** can cause unfairness in the system, pointing out that *"biases may be present due to the human programmers who developed the algorithm"* (p18). Finally, only three referred to the **Dataset** used to train the system. One explained *"gender bias based on the data input showing that one gender is more successful than the other"* (p3).

| | AG | UN | AF | FP | DO | TR | Friedman's Test ($x^2(2)$) | p | Mann-Whitney | p | M (Blacks) | M (Whites) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AG | 1 | | | | | | 86.575 | <0.001 | 1888.5 | 0.050 | 5.000 | 4.000 |
| UN | 0.479 | 1 | | | | | 117.580 | <0.001 | 1816.5 | 0.021 | 5.000 | 4.000 |
| AP | 0.764 | 0.578 | 1 | | | | 306.553 | <0.001 | 1605.5 | 0.002 | 4.000 | 3.000 |
| FP | 0.776 | 0.526 | 0.816 | 1 | | | 115.868 | <0.001 | 1648.5 | 0.003 | 5.000 | 3.000 |
| DO | 0.786 | 0.408 | 0.670 | 0.790 | 1 | | 103.507 | <0.001 | 1720 | 0.007 | 4.000 | 3.000 |
| TR | 0.603 | 0.385 | 0.645 | 0.700 | 0.678 | 1 | 90.744 | <0.001 | 1538.5 | <0.001 | 4.000 | 2.000 |

Table 5: **Left:**Pearson Correlations for the six constructs of justice. All correlations are significant at p <0.001 level.**Centre:** Friedman's Test results when comparing the six constructs change in the same scenario, when different decisions are provided. **Right:** Mann-Whitney U Test, for differences between the two ethnicity groups for Scenario 3, Case A.

| Category | Description | # | DO | OF | MR | EN | TD | DC | HI | C | ET | DG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Biases/Discrimination | *existance of biases/discrimination* | 51 | 16 | 13 | 6 | 9 | 7 | 8 | 4 | 3 | 0 | 3 |
| Decision/Outcome | *the quality of the decision or outcome* | 50 | | 15 | 9 | 11 | 8 | 6 | 1 | 5 | 1 | 4 |
| Objective Factors | *relevance/appropriateness of factors* | 34 | | | 11 | 5 | 2 | 4 | 1 | 5 | 0 | 1 |
| Methods/Rules | *appropriate procedures* | 27 | | | | 6 | 8 | 0 | 4 | 1 | 0 | 0 |
| Moral/Ethics/Norms | *consideration of ethics or social norms* | 26 | | | | | 8 | 0 | 4 | 1 | 0 | 0 |
| Training Data | *the impact of the dataset used* | 22 | | | | | | 4 | 4 | 1 | 0 | 3 |
| Demographics | *sensitive attributes (e.g. gender, race)* | 19 | | | | | | | 4 | 1 | 0 | 0 |
| Human Intervention | *the impact of humans* | 13 | | | | | | | | 2 | 0 | 3 |
| Context | *consideration of different scenarios* | 11 | | | | | | | | | 2 | 0 |
| Transparency | *the algorithm/output is transparent* | 4 | | | | | | | | | | 0 |
| Disadvantaged Groups | *the impact on disadvantaged groups* | 4 | | | | | | | | | | |
| Other | *[falls outside of established themes]* | 25 | | | | | | | | | | |

Table 6: Themes for Q2: (a)name, description (full description in (Kleanthous et al. 2022)) and frequency, (b) co-occurrences

## Algorithmic Accountability

When asked who would be held accountable (Q8) the majority (76%) agreed that "the company using the system" (S3), 67.3% indicated that "the company that owns the system" (S2), 56.7% opted for "the team that developed the system" (S1), and 52.7% believed "the hiring manager who uses the system" (S4) would be held accountable. Only 2% agreed that "neither the system nor the team" (S5) should be accountable. In their explanations (Q9), those who selected S3 remarked that *"the company is ultimately responsible for [their] hiring practices "* (p61) and *"[for] accept[ing] any biased hiring system"* (p26). Those who selected S2 argued that *"the company that owns the system holds more responsibility"* (p132).In contrast, those who chose S1 highlighted that *"AI can be trained to be racist if developers feed it such"*(p48) and *"[it] does what it was designed to do, so the people designing it are liable"* (p76). Finally, those choosing S4, mentioned that *" [the] hiring manager should know what system they are using* (p32) *"[and] report such an unfair system"* (p27), thus he/she would be accountable.

82 selected S2 and S3 noting that *"Companies hold an implied fiduciary duty of trust to their Employees (prospective or confirmed) to act according to and in line with the Employment and Labour laws"* (p114).14 out of 82 opted only for S2 and S3 arguing that *"the team which developed the system and the hiring manager wouldn't be held accountable because they just did what the company told them to do, so only the company is accountable"* (p71). An important view was revealed by those (68) selecting S3 and S4. They argued that *"it's always the one who uses it, who should be responsible for the outcome"* (p56), since *"the company and hiring manager should know what system they are using"* (p32). 11 selected S2, S3, and S4 arguing that *"the designing team will only set the parameters required by the company, anyone who uses it to make a decision or owns it should be held accountable for the discrimination that happens after using the system"* (p63). Others (42) felt like *"everyone involved must be held accountable"*.

## Discussion

When examining the relationship between the six constructs of fairness, we observed that our participants responded differently to the outcome of the DSS and its impact on the individual in the scenario, compared to the CS affiliates (Kasinidou et al. 2021a). Unlike CS participants, who largely aligned with the system's decision, those in our study disagreed with the decision presented in Scenario 2 and felt that the individual did not deserve the resulting outcome. This is an indication that people with different backgrounds perceive fairness constructs differently. Participants generally viewed the factors used in the DM process in Scenarios 1 and 2 as appropriate. However, they still perceived the overall process as unfair, noting that additional factors should have been taken into account. This contrasts with the findings of Kleanthous et al., who reported that participants not only judged the decision as unfair, particularly in Scenario 2, but also deemed the factors themselves (e.g., gender and age) as inappropriate for consideration. The legal participants also mentioned gender and age as factors that "can cause problems" when taken into account in the DSS. Interestingly, while Kleanthous et al. noted that their participants explicitly stated it is illegal to use gender and race in algorithmic decision-making, none of our participants made any such reference. Depending on the context, legal affili-

| Category | Description | # |
|---|---|---|
| Sensitive Attributes | *use of factors such as gender, race, age* | 116 |
| System/Model | *the procedures followed by the model* | 9 |
| *Lack of appropriate factors | *not considering the most appropriate parameters* | 6 |
| *No Human Judge | *humans can consider more parameters* | 5 |
| Human Influence | *the impact of humans on the system* | 5 |
| Dataset | *the data used for training the model or as input* | 3 |
| Other | *[falls outside of the established themes]* | 23 |

Table 7: Thematic analysis for Q3: name, description and frequency. Themes with * are the themes added in this study

ates have different views on whether the factors considered are appropriate, even when they do agree with the decision. In general, we can see from the emerging themes that they focused on different aspects of fairness compared to the participants in (Kasinidou et al. 2021b; Kleanthous et al. 2022).

In Scenario 3, our participants perceived Case A as the most just, compared to the participants in (Kasinidou et al. 2021a), where Case B (proportional outcome) was selected. In the themes that emerged from their free-text responses, "Unfair" was mentioned in Cases B and C but not in Case A. However, they also mentioned the need for more information on the parameters used and the importance of certain factors in DM. Furthermore, when asked whether the information provided was sufficient, a new theme emerged across all three scenarios: participants described the system's decision as "unfair" or "discriminatory." This differs with the findings of (Kleanthous et al. 2022), where CS affiliates did not raise such concerns, as their focus remained primarily on the algorithms themselves.

Our results indicate that the participants lack the necessary knowledge in topics related to FATE in algorithmic DSS. This result supports prior work (Yu and Alì 2019), indicating that while law schools acknowledge the increasing use of DSS in legal practice, there are only a handful of resources available around the world. Previous studies highlighted that even short seminars can make a difference in understanding and the attitude of practitioners (Kasinidou et al. 2021b; Pierson 2017). Thus, an important output of this work is to stress the need for new programs that will educate future legal practitioners on algorithmic DSS in the context of legal practice.

**Defining Algorithmic Fairness:** While some participants, as expected (Colquitt and Rodell 2015; Pfeiffer et al. 2023), struggled to provide a comprehensive definition, the majority described algorithmic fairness as the absence of bias or discrimination in DM. Their responses likely reflect familiarity with high-profile cases of algorithmic bias, such as COMPAS[3], where fairness is framed through a legal lens emphasizing non-discrimination (Pfeiffer et al. 2023; Pessach and Shmueli 2023). However, only a few participants acknowledged the role of the specific features used by the system or the underlying algorithmic logic in shaping fairness. This contrasts with prior work by (Kleanthous et al. 2022), where CS students were more attuned to the influence of input factors and algorithmic processes on fairness. These findings suggest that perceptions of algorithmic fairness are strongly shaped by domain-specific knowledge and training. The limited awareness of how these systems function highlights the need for greater education and exposure to AI concepts among legal affiliates engaging with DSSs.

**Algorithmic Accountability:** When participants were asked to comment on accountability, the majority agreed that both "the company that is using the system" and "the company that owns the system" should be held accountable. This finding is in contrast to earlier work (Kleanthous et al. 2022) where participants believed that the team that developed the system should be held accountable. It is a confirmation that law students and young legal professionals have a different perspective on this matter than other groups. It further stresses the need to establish a common ground between developers, owners, and those who will be judging the behavior of these systems.

**Limitations.** A key limitation of this study is the participant pool, crowdworkers with legal backgrounds, whose perspectives on fairness in DM, while valuable, may not generalize to broader populations. The thematic analysis identifies prevalent themes but not the intensity or nuance of participants' views. Though many noted missing factors or procedural concerns, their feedback was often vague and lacked specific suggestions, underscoring the difficulty of obtaining actionable input from non-technical stakeholders.

## Conclusion

Algorithmic DSS are widely used across industries, yet limited awareness of their impacts can mask serious societal risks. Legal affiliates are crucial in shaping regulations and pursuing legal remedies when harm occurs, making their perspectives essential. This study shows their views often diverge from those of computer scientists and highlights a shared gap in AI ethics training across both fields. These findings support the need for targeted education and policies to better align stakeholder perspectives.

## Acknowledgments

This project is partially funded by the EU's Horizon Europe Programme under agreement No. 101214000 (TRUMAN) and the Cyprus Research and Innovation Foundation Programme under agreement No. BRIDGE2HORIZON/0823E/0203 (PINNACLE).

[3]https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

## References


Allen, A. 2016. The 'three black teenagers' search shows it is society, not Google, that is racist. *The Guardian*, 10.

Angerschmid, A.; Zhou, J.; Theuermann, K.; Chen, F.; and Holzinger, A. 2022. Fairness and Explanation in AI-Informed Decision Making. *Machine Learning and Knowledge Extraction*, 4(2): 556–579.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. ProPublica. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica, 23 Mai 2016.

Araujo, T.; Helberger, N.; Kruikemeier, S.; and De Vreese, C. H. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & society*, 35: 611–623.

Ashong Elliot, M. A.; and Arthur, R. 2020. Organizational Justice: Does "IT" Matter? Empirical Analysis of the Influence of Information Technology on Employee Justice Perceptions. In Nunes, I. L., ed., *Advances in Human Factors and Systems Interaction*, 83–89. Cham: Springer International Publishing. ISBN 978-3-030-51369-6.

Aysolmaz, B.; Müller, R.; and Meacham, D. 2023. The public perceptions of algorithmic decision-making systems: Results from a large-scale survey. *Telematics and Informatics*, 79: 101954.

Barabas, C.; Dinakar, K.; and Doyle, C. 2019. The problems with risk assessment tools. *The New York Times*, 7–9.

Barlas, P.; Kyriakou, K.; Guest, O.; Kleanthous, S.; and Otterbacher, J. 2021. To "See" is to Stereotype: Image Tagging Algorithms, Gender Recognition, and the Accuracy-Fairness Trade-Off. *Proc. ACM Hum.-Comput. Interact.*, 4(CSCW3).

Barysė, D. 2022. Do we need more technologies in courts? Mapping concerns for legal technologies in courts. *Mapping Concerns for Legal Technologies in Courts (September 6, 2022)*.

Barysė, D.; and Sarel, R. 2024. Algorithms in the court: does it matter which part of the judicial decision-making is automated? *Artificial intelligence and law*, 1–30.

Benatti, R.; Severi, F.; Avila, S.; and Colombini, E. L. 2024. Gender Bias Detection in Court Decisions: A Brazilian Case Study. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 746–763. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Binns, R.; Van Kleek, M.; Veale, M.; Lyngs, U.; Zhao, J.; and Shadbolt, N. 2018. "It's Reducing a Human Being to a Percentage": Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206.

Blandin, J.; and Kash, I. A. 2024. Learning Fairness from Demonstrations via Inverse Reinforcement Learning. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 51–61. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Bogiatzis-Gibbons, D. J. 2024. Beyond Individual Accountability: (Re-)Asserting Democratic Control of AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, 74–84. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704505.

Bozdag, E. 2013. Bias in algorithmic filtering and personalization. *Ethics and information technology*, 15(3): 209–227.

Brown, A.; Chouldechova, A.; Putnam-Hornstein, E.; Tobin, A.; and Vaithianathan, R. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Chandra, R.; and Sanjaya, K. 2023. Punishing the Unpunishable: A Liability Framework for Artificial Intelligence Systems. In *International Conference on Digital Technologies and Applications*, 55–64. Springer.

Chen, I. Y.; Johansson, F. D.; and Sontag, D. 2018. Why is My Classifier Discriminatory? In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, 3543–3554. Red Hook, NY, USA: Curran Associates Inc.

Chouldechova, A. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2): 153–163.

Colquitt, J. A.; and Rodell, J. B. 2015. Measuring justice and fairness.

Diakopoulos, N. 2016. Accountability in algorithmic decision making. *Communications of the ACM*, 59(2): 56–62.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through Awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, 214–226. New York, NY, USA: Association for Computing Machinery. ISBN 9781450311151.

Eslami, M.; Vaccaro, K.; Lee, M. K.; Elazari Bar On, A.; Gilbert, E.; and Karahalios, K. 2019. User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Green, B.; and Hu, L. 2018. The myth in the methodology: Towards a recontextualization of fairness in machine learning. In *Proceedings of the machine learning: the debates workshop*. Stockholm, Sweden.

Greenstein, S. 2022. Preserving the rule of law in the era of artificial intelligence (AI). *Artificial Intelligence and Law*, 30(3): 291–323.

Grgić-Hlača, N.; Zafar, M. B.; Gummadi, K. P.; and Weller, A. 2018. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 51–60. Palo Alto: CA: AAAI.

Grgić-Hlača, N.; Redmiles, E. M.; Gummadi, K. P.; and Weller, A. 2018. Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, 903–912. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee. ISBN 9781450356398.

Harasta, J.; Novotná, T.; and Savelka, J. 2024. It Cannot Be Right If It Was Written by Ai: On Lawyers? Preferences of Documents Perceived as Authored by an Llm Vs a Human. *Artificial Intelligence and Law*, 1–38.

Harrison, G.; Hanson, J.; Jacinto, C.; Ramirez, J.; and Ur, B. 2020. An Empirical Study on the Perceived Fairness of Realistic, Imperfect Machine Learning Models. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 392–402. New York, NY, USA: Association for Computing Machinery. ISBN 9781450369367.

Helberger, N.; Araujo, T.; and de Vreese, C. H. 2020. Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making. *Computer Law & Security Review*, 39: 105456.

Hermstrüwer, Y.; and Langenbach, P. 2023. Fair governance with humans and machines. *Psychology, Public Policy, and Law*.

Holstein, K.; Wortman Vaughan, J.; Daumé, H.; Dudik, M.; and Wallach, H. 2019. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359702.

Kasinidou, M.; Kleanthous, S.; Barlas, P.; and Otterbacher, J. 2021a. I Agree with the Decision, but They Didn't Deserve This: Future Developers' Perception of Fairness in Algorithmic Decisions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 690–700. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Kasinidou, M.; Kleanthous, S.; Orphanou, K.; and Otterbacher, J. 2021b. Educating Computer Science Students about Algorithmic Fairness, Accountability, Transparency and Ethics. In *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, ITiCSE '21, 484–490. New York, NY, USA: Association for Computing Machinery. ISBN 9781450382144.

Kleanthous, S.; Kasinidou, M.; Barlas, P.; and Otterbacher, J. 2022. Perception of fairness in algorithmic decisions: future developers' perspective. *Patterns*, 3(1).

Kontiainen, L.; Koulu, R.; and Sankari, S. 2022. Research agenda for algorithmic fairness studies: Access to justice lessons for interdisciplinary research. *Frontiers in Artificial Intelligence*, 5.

Kyriakou, K.; Kleanthous, S.; Otterbacher, J.; and Papadopoulos, G. A. 2020. Emotion-Based Stereotypes in Image Analysis Services. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20 Adjunct, 252–259. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379502.

Lahoti, P.; Gummadi, K. P.; and Weikum, G. 2019a. ifair: Learning individually fair data representations for algorithmic decision making. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 1334–1345. IEEE.

Lahoti, P.; Gummadi, K. P.; and Weikum, G. 2019b. Operationalizing individual fairness with pairwise fair representations. *Proceedings of the VLDB Endowment*, 13(4): 506–518.

Lee, M. K. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1): 2053951718756684.

Lee, M. K.; and Baykal, S. 2017. Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, 1035–1048. New York, NY, USA: Association for Computing Machinery. ISBN 9781450343350.

Leventhal, G. 1980. What Should Be Done With Equity Theory?'Social Exchange: Advances.

Lindquist, S.; and Cross, F. 2012. Stability, Predictability and The Rule of Law: Stare Decisis As Reciprocity Norm. Citeseerx.

Longoni, C.; Bonezzi, A.; and Morewedge, C. K. 2019. Resistance to Medical Artificial Intelligence. *Journal of Consumer Research*, 46(4): 629–650.

Malek, M. A. 2022. Criminal courts' artificial intelligence: the way it reinforces bias and discrimination. *AI and Ethics*, 2(1): 233–245.

Marcinkowski, F.; Kieslich, K.; Starke, C.; and Lünich, M. 2020. Implications of AI (Un-)Fairness in Higher Education Admissions: The Effects of Perceived AI (Un-)Fairness on Exit, Voice and Organizational Reputation. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 122–130. New York, NY, USA: Association for Computing Machinery. ISBN 9781450369367.

McGregor, L.; Murray, D.; and Ng, V. 2019. INTERNATIONAL HUMAN RIGHTS LAW AS A FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY. *International Comparative Law Quarterly*, 68(2): 309–343.

Mujtaba, D. F.; and Mahapatra, N. R. 2019. Ethical considerations in ai-based recruitment. In *2019 IEEE International Symposium on Technology and Society (ISTAS)*, 1–7. IEEE.

Nagtegaal, R. 2021. The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly*, 38(1): 101536.

Nakao, Y.; Strappelli, L.; Stumpf, S.; Naseer, A.; Regoli, D.; and Del Gamba, G. 2023. Towards Responsible AI: A Design Space Exploration of Human-Centered Artificial

Intelligence User Interfaces to Investigate Fairness. *International Journal of Human–Computer Interaction*, 39(9): 1762–1788.

Otterbacher, J.; Bates, J.; and Clough, P. 2017. Competent Men and Warm Women: Gender Stereotypes and Backlash in Image Search Results. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 6620–6631. New York, NY, USA: Association for Computing Machinery. ISBN 9781450346559.

Pessach, D.; and Shmueli, E. 2023. Algorithmic fairness. In *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, 867–886. Springer.

Pfeiffer, J.; Gutschow, J.; Haas, C.; Möslein, F.; Maspfuhl, O.; Borgers, F.; and Alpsancar, S. 2023. Algorithmic Fairness in AI: An Interdisciplinary View. *Business & Information Systems Engineering*, 65(2): 209–222.

Pierson, E. 2017. Demographics and discussion influence views on algorithmic fairness. arXiv:1712.09124.

Rader, E.; Cotter, K.; and Cho, J. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206.

Rader, E.; and Gray, R. 2015. Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 173–182. New York, NY, USA: Association for Computing Machinery. ISBN 9781450331456.

Rahwan, I.; Cebrian, M.; Obradovich, N.; Bongard, J.; Bonnefon, J.-F.; Breazeal, C.; Crandall, J. W.; Christakis, N. A.; Couzin, I. D.; Jackson, M. O.; et al. 2019. Machine behaviour. *Nature*, 568(7753): 477–486.

Ramesh, D.; Kameswaran, V.; Wang, D.; and Sambasivan, N. 2022. How Platform-User Power Relations Shape Algorithmic Accountability: A Case Study of Instant Loan Platforms and Financially Stressed Users in India. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, 1917–1928. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393522.

Said, G.; Azamat, K.; Ravshan, S.; and Bokhadir, A. 2023. Adapting Legal Systems to the Development of Artificial Intelligence: Solving the Global Problem of AI in Judicial Processes. *International Journal of Cyber Law*, 1(4).

Saxena, N. A.; Huang, K.; DeFilippis, E.; Radanovic, G.; Parkes, D. C.; and Liu, Y. 2019. How Do Fairness Definitions Fare? Examining Public Attitudes Towards Algorithmic Definitions of Fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, 99–106. New York, NY, USA: Association for Computing Machinery. ISBN 9781450363242.

Srivastava, M.; Heidari, H.; and Krause, A. 2019. Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, 2459–2468. New York, NY, USA: Association for Computing Machinery. ISBN 9781450362016.

Thorson, K.; Cotter, K.; Medeiros, M.; and Pak, C. 2019. Algorithmic inference, political interest, and exposure to news and politics on Facebook. *Information, Communication & Society*, 0(0): 1–18.

Tran, T. N. T.; Atas, M.; Felfernig, A.; Le, V. M.; Samer, R.; and Stettinger, M. 2019. Towards Social Choice-Based Explanations in Group Recommender Systems. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '19, 13–21. New York, NY, USA: Association for Computing Machinery. ISBN 9781450360210.

Tsai, C.-H.; and Brusilovsky, P. 2019. Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '19, 22–30. New York, NY, USA: Association for Computing Machinery. ISBN 9781450360210.

van Berkel, N.; Sarsenbayeva, Z.; and Goncalves, J. 2023. The methodology of studying fairness perceptions in Artificial Intelligence: Contrasting CHI and FAccT. *International Journal of Human-Computer Studies*, 170: 102954.

Veale, M.; Van Kleek, M.; and Binns, R. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. CHI '18, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206.

Wang, A. 2018. Procedural Justice and Risk-Assessment Algorithms. *Available at SSRN 3170136*.

Wang, R.; Harper, F. M.; and Zhu, H. 2020. Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Woodruff, A.; Fox, S. E.; Rousso-Schindler, S.; and Warshaw, J. 2018. A Qualitative Exploration of Perceptions of Algorithmic Fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356206.

Yalcin, G.; Themeli, E.; Stamhuis, E.; Philipsen, S.; and Puntoni, S. 2023. Perceptions of justice by algorithms. *Artificial Intelligence and Law*, 31(2): 269–292.

Yu, R.; and Alì, G. S. 2019. What's inside the black box? AI challenges for lawyers and researchers. *Legal Information Management*, 19(1): 2–13.

Zemel, R.; Wu, Y.; Swersky, K.; Pitassi, T.; and Dwork, C. 2013. Learning Fair Representations. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, III–325–III–333. JMLR.org.