Information Systems Frontiers
https://doi.org/10.1007/s10796-021-10223-8

AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context

Sarah Bankins[1] · Paul Formosa[2] · Yannick Griep[3] · Deborah Richards[4]

Accepted: 2 November 2021

Abstract
Using artificial intelligence (AI) to make decisions in human resource management (HRM) raises questions of how fair employees perceive these decisions to be and whether they experience respectful treatment (i.e., interactional justice). In this experimental survey study with open-ended qualitative questions, we examine decision making in six HRM functions and manipulate the decision maker (AI or human) and decision valence (positive or negative) to determine their impact on individuals' experiences of interactional justice, trust, dehumanization, and perceptions of decision-maker role appropriateness. In terms of decision makers, the use of human decision makers over AIs generally resulted in better perceptions of respectful treatment. In terms of decision valence, people experiencing positive over negative decisions generally resulted in better perceptions of respectful treatment. In instances where these cases conflict, on some indicators people preferred positive AI decisions over negative human decisions. Qualitative responses show how people identify justice concerns with both AI and human decision making. We outline implications for theory, practice, and future research.

Keywords Artificial intelligence · Human resource management · Algorithmic management · Ethical AI · Artificial intelligence at work · Interactional justice

1 Introduction

Artificially intelligent (AI) systems are increasingly being used in the domain of human resource management (HRM). In a process referred to as "algorithmic management", AI systems are being used to assess applicants in the recruitment and selection process (Marr, 2018), allocate work (Lee et al., 2015), provide training recommendations (Guenole & Feinzig, 2018), and terminate workers' employment (Kellogg et al., 2020). Given that HRM decisions can directly impact the livelihoods of humans, the use of AI in this domain raises ethical questions about how *fair* or *just* employees believe these decisions to be. In an organizational context such perceptions of fairness fall under the notion of organizational justice, defined as the "perceived adherence to rules that reflect appropriateness in decision contexts" (Colquitt & Rodell, 2015, p. 188). Justice perceptions are important to examine because if employees believe organizational decision making is fair, they are more likely to accept the decision, remain satisfied in their jobs, and even increase their level of effort (Lind, 2001). However, if employees perceive reduced organizational justice, for example because

Sarah Bankins and Paul Formosa contributed equally to the manuscript and share first authorship

✉ Sarah Bankins
sarah.bankins@mq.edu.au

Paul Formosa
paul.formosa@mq.edu.au

Yannick Griep
y.griep@psych.ru.nl

Deborah Richards
deborah.richards@mq.edu.au

[1] Department of Management, Macquarie Business School, Macquarie University, North Ryde campus, Sydney, NSW 2109, Australia

[2] Department of Philosophy, Faculty of Arts, Macquarie University, North Ryde Campus, Sydney, NSW 2109, Australia

[3] Behavioural Science Institute, Radboud University, Postbus 9104, 6500 HE Nijmegenn, Netherlands

[4] Department of Computing, Faculty of Science and Engineering, Macquarie University, North Ryde campus, Sydney, NSW 2109, Australia

Published online: 02 February 2022

they do not trust AI decision making or find it inappropriate or dehumanizing, they may experience reduced effort, lower job satisfaction, lower organizational commitment, and higher likelihood of turnover (Lind, 2001).

However, existing research into justice perceptions of AI decision making in the workplace largely focuses on procedural (how fair and reasonable the *procedures* used to make a decision are) and distributive (how fair the *outcomes* of a decision are, such as resource allocation) forms of justice (e.g., Lee et al., 2019; Tambe et al., 2019; see Robert et al.,'s, 2020 review). However, empirical work suggests that in decision making individuals seek a "human touch" and chafe against "being reduced to a percentage" when AI is responsible for making decisions that impact them (Binns et al., 2018, p. 1). This implicates another comparatively neglected justice perception, *interactional justice*, as a critical aspect for better understanding the conditions under which employees perceive AI decision making to be fair or unfair. Interactional justice refers to individuals' beliefs that they have been treated with dignity and respect (Bies, 2001), such as when decision makers demonstrate sensitivity and empathy and adequately explain a decision (Skarlicki & Folger, 1997). Poor interactional justice perceptions can result in employees' experiencing dehumanization (such as mechanistic dehumanization, or feelings of being treated as a robot or inanimate object) and feelings of indignity and disrespect (Bies, 2001; Haslam, 2006). These feelings have been shown to have significant adverse effects not only on employees' performance, but also on their wider health and well-being (Christoff, 2014; Lucas, 2015).

Given the advancing involvement of AI when managing people at work, it is critical to examine how people respond to such systems making decisions about them in this context and whether they construe these decisions as fair. This is particularly important as the role of AI in employment decisions is increasingly recognized as a high-risk context for the use of these technologies (e.g., European Commission, 2021). This regulatory background also couples with broad community concerns regarding the *trustworthiness* of AI technologies (Lockey et al., 2020), their *appropriateness* in some decision-making contexts (Lee, 2018), and that their use can be *dehumanizing* (Binns et al., 2018). Taken together, these issues make it important and timely to assess how the use of AI for HRM decision making influences individuals' feelings of interactional justice, given the potentially negative effects of experiences of injustice upon individuals' well-being, and to inform practitioners and policy-makers on appropriate deployment and regulatory settings to guide the technology's use. In this context, the research question driving this study is: *How does AI decision making, compared to human decision making, impact individuals' perceptions of interactional (in)justice?* In examining this question, we also assess the importance of trust, dehumanization, and role appropriateness, given their links to interactional justice as identified above.

Using an experimental survey study, we construct decision-making scenarios across six HRM functions that reflect the ways AI technologies currently impact workers (aligned to Kellogg et al., 2020) through determining: recruitment and selection outcomes; training recommendations; performance management; work allocation; promotion outcomes; and firing. We vary the decision maker (AI or human) and the decision valence (positive or negative for the worker) and examine the influence on individuals' interactional justice perceptions. We apply three complementary analytical lenses to the data through undertaking statistical analyses (MANOVA), machine learning, and qualitative thematic coding of open-ended survey responses. This generates insights into how both the decision maker (AI versus human) and decision valence (positive versus negative) enhance or diminish interactional justice perceptions and, more broadly, affords insights into how individuals construe the positive and negative implications of AI versus human decision making for HRM functions.

2 Literature Review

Organizational justice is traditionally comprised of different forms of justice: procedural; distributive; and interactional (Cropanzano et al., 2001). Unlike other justice perceptions that focus on the reasonableness of decision-making procedures (procedural) and the fairness of decision outcomes (distributive), interactional justice focuses on the "interpersonal treatment" experienced by individuals in the decision-making process (Erdogan, 2002, p. 557). Bies and Moag (1986) suggest that positive interpersonal treatment involves being treated with respect and dignity and being offered truthful and justifiable explanations for outcomes received. To address the overarching research question presented in the introduction, the study focuses on two aspects of decision making that we argue will impact interactional justice perceptions in a HRM context: (1) the type of decision maker; and (2) the decision valence. The following subsections elaborate on these two aspects.

2.1 Decision Maker: AI versus Human

While more research is required to assess how human versus AI decision making influences people's views of a decision taken (for existing work see Lee, 2018), there is evidence that who or what is making the decision will influence perceptions of interpersonal treatment. For example, HRM decisions can be sensitive in nature, such as performance management or hiring and firing decisions, and these decisions may have significant implications for workers'

well-being. Existing research suggests that in such cases workers expect human judgment and intuition to be used (Lee, 2018; see Binns, 2020 for normative arguments) and human interaction to be provided (Binns et al., 2018; Lee et al., 2019), and that algorithms can be viewed as failing to account for "human abilities, emotion, and motivation" (Lee et al., 2015, p. 5) and do not have the requisite capabilities to make HRM-based decisions (Lee, 2018).

This research shows how the use of AI for decision making can challenge key tenets underpinning experiences of interactional justice. For example, because people form expectations regarding the roles individuals play in an organization (Hamilton, 1978), the more appropriate the decision maker is perceived to be (*role appropriateness*) the more likely that Bies and Moag's (1986) dimensions of propriety, respect, dignity, and appropriate justifications (*interactional justice*) will be met. However, the work of Lee and colleagues shows that workers find it challenging to characterize AI as an *appropriate decision maker* in a HRM context, given the limitations they identify in its decision-making processes. These concerns, combined with the lack of human interaction often attendant in AI decision making, can lead workers to think that they are "being reduced to a percentage" (Binns et al., 2018, p. 1; Lee et al., 2019, p. 16; Lee, 2018); which constitutes a set of feelings often captured under the umbrella term of *dehumanization.* Finally, some scholars have argued that workers perceive lower levels of *trust* in decisions when an AI, as opposed to a human, is the decision maker (for examples see Karunakaran, 2018; Ticona & Mateescu, 2018). While there is evidence that people can engage in "algorithm appreciation", or a preference for the determinations of algorithms (Logg et al., 2019, p. 90), this is often found for decisions that are relatively objective, may not directly impact the individual, and occur in contexts quite different to HRM decisions. We thus hypothesize that, in a HRM context:

> *H1: When an AI is the decision maker, people will experience lower interactional justice (H1a), lower perceptions of decision-maker role appropriateness (H1b), lower levels of trust (H1c) and higher levels of dehumanization (H1d) compared to when the decision maker is human.*

2.2 Decision Valence: Positive versus Negative

Beyond attributes of the decision maker, research shows that people evaluate the quality of decisions based on the valence (either positive or negative) of the outcome of those decisions, sometimes regardless of the process taken to reach them (Fischhoff, 1975). Labelled "outcome bias", and the related "hindsight bias", evaluations of decisions often occur after the fact and therefore incorporate outcome information, despite this often not being the optimal indicator of decision quality (Lipshitz, 1989). Outcome bias exists when individuals view decisions with positive outcomes more favorably and decisions with negative outcomes more negatively. Empirically, Lipshitz (1989) shows that when decisions are viewed as successful (or generating positive outcomes), then those decisions are evaluated as more justified and stemming from superior decision making. Outcome bias can also be amplified when individuals have little relevant information upon which to base their evaluations of decisions (Baron & Hershey, 1988). However, how outcome bias manifests in the context of AI as a decision maker has not yet been explored. As the use of AI for decision making has well-documented benefits and costs (see Colson, 2019; Shrestha et al., 2019), and the above-mentioned research shows people are cognizant of these in the decision-making processes associated with HRM, we also assess the impact of the valence of a decision outcome (either positive or negative) on individuals' assessments of *interactional justice*, *trust, dehumanization*, and perceptions of decision-maker *role appropriateness.* Based on the above theoretical work and accompanying empirical support that people view positive valence outcomes as more justified and favorable compared to their negative counterparts, we thus hypothesize that in a HRM context:

> *H2: Regardless of decision maker, when an individual receives a positive valence outcome they will experience higher interactional justice (H2a), higher perceptions of decision-maker role appropriateness (H2b), higher levels of trust (H2c) and lower levels of dehumanization (H2d) compared to when they receive a negative valence outcome.*

H1 and H2 conflict in the case where an AI decides a positive outcome and a human decides a negative outcome, as H1 leads us to believe that people will prefer the human decision because of who the decision maker is, while H2 leads us to believe that people will prefer the AI decision maker because the decision is positive. This is an interesting case because it examines which aspect is more important in terms of interactional justice perceptions: *who* makes the decision or the *valence* of the decision. Given the competing arguments in this case, our assessment of it is exploratory (termed 'exploratory case' in our results section).

3 Methods

3.1 Research Design

Experimental survey studies allow researchers to examine how individuals respond to different scenarios by altering factors that are believed to influence individuals' judgments

Table 1 Vignette Overview and Groupings by Decision Maker and Decision Valence (Outcome)

Groups	Specific vignettes
AI- Group: AI is decision maker and decides negative (i.e., a benefit is not recommended)	• AI decision maker / training not recommended • AI decision maker / work allocation not recommended • AI decision maker / promotion not recommended • AI decision maker / hiring not recommended • AI decision maker / firing recommended • AI decision maker / performance management not recommended (no bonus provided)
AI + Group: AI is decision maker and decides positive (i.e., a benefit is recommended)	• AI decision maker / training recommended • AI decision maker / work allocation recommended • AI decision maker / promotion recommended • AI decision maker / hiring recommended • AI decision maker / firing not recommended • AI decision maker / performance management recommended (bonus provided)
H- Group: Human is decision maker and decides negative	• Human decision maker / training not recommended • Human decision maker / work allocation not recommended • Human decision maker / promotion not recommended • Human decision maker / hiring not recommended • Human decision maker / firing recommended • Human decision maker / performance management not recommended (no bonus provided)
H + Group: Human is decision maker and decides positive	• Human decision maker / training recommended • Human decision maker / work allocation recommended • Human decision maker / promotion recommended • Human decision maker / hiring recommended • Human decision maker / firing not recommended • Human decision maker / performance management recommended (bonus provided)

of a phenomena (Wallander, 2009). We utilized a 2 × 2 experimental survey design, manipulating the decision maker (AI/human) and the decision valence (positive/negative). We then applied these manipulations to six HRM functions where AI is currently used in decision making, per Kellogg et al.'s work (2020). In their paper, they argue that algorithms can shape employees' experiences along '6 Rs': by directing workers (through **R**estricting and **R**ecommending information or courses of action); evaluating workers (through **R**ecording and **R**ating behaviors); and disciplining workers (through **R**eplacing or **R**ewarding them). Our six HRM functions, and their alignment to these '6 Rs' are: recruitment and selection (rating); training (recommending); performance management (rewarding); work allocation (restricting, through directing workers); firing (replacing); and promotion (rating). While in practice there will be overlap between these categories (e.g., promotion can represent both rating and recording), we aimed to cover a range of HRM functions in which AI is currently being used to make decisions. This approach led to a 2 × 2 × 6 design, with a total of 24 experimental vignettes. For all our analyses we grouped the vignettes into four groups per our 2 × 2 manipulations (see Table 1): AI decides negative (AI-); AI decides positive (AI+); human decides negative (H-); and human decides positive (H+).[1]

3.2 Materials and Procedures

The structure of each vignette was consistent, approximately the same length, and reflected the following information: (1) what the vignette is focused on (i.e., the focal HRM function and decision maker); (2) why the situation is important to the individual (i.e., to ensure respondents understand that they value what is being provided in the scenario); (3) the relevant employee data gathered—each vignette identified four pieces of data relevant to the scenario that included three objectively derived pieces of data (e.g., number of sales) and one subjectively derived piece of data (e.g., positive contribution to team dynamics)—to ensure the information used in the decision was clear and consistent across human and AI versions; and (4) the decision made (i.e., the valence of the decision outcome). In reading each vignette, participants were asked to place themselves as employees in a fictitious company

[1] To confirm this grouping as appropriate, we also conducted ANOVAs grouping the vignettes by the six HRM functions and found no significant differences between groups on any of our variables of interest (results available on request from the authors).

('Triton', a professional services firm) and were provided with some detail on their role and duties in this company. To support the external validity of the vignettes, and following best practice recommendations (Aguinis & Bradley, 2014), each vignette was reviewed by several organizational behavior academics and two human resources practitioners with experience in the use of AI in workplaces (in the financial, educational, and tourism sectors).

For the individuals who received vignettes with *AI as the decision maker* (termed an 'AI algorithm'), the explanation of 'AI algorithm' was: "The term (AI algorithm) means a series of rules or procedures an artificial intelligence system uses to decide what to do. An AI algorithm may, for example, make inferences and predictions based on data". For the individuals who received vignettes with a *human manager as the decision maker*, the explanation of this term was: "The term (manager) is used to refer to the person who is your direct line manager, or the equivalent person. That is, the person who is involved in the day-to-day and direct management of you as an employee of the company". An example vignette is below, reflecting the training HRM function with the manipulated components identified in square brackets:

> *Triton is offering its employees in the customer service department the opportunity to undertake a multi-day training course. However, there are a limited number of places available and this training course will only be offered once. You really want to attend the training because it will help you upskill and you believe it will help with your future career prospects. Triton gathers data on employees' current skills, how they applied skills learnt in previous training to their current job, their training attendance history, and the projected future skills needed in their roles. [Your manager/An AI algorithm] has analyzed this employee training data and [you are recommended for the training/you are not recommended for the training].*

Participants were recruited through CloudResearch, a data services provider that draws on the working adult North American population. Online panels are a reliable source for accessing diverse samples (e.g., Landers & Behrend, 2015), with the quality of data not being substantially different compared to a non-paid random sample (e.g., Behrend et al., 2011), especially when researchers embed (as we did) attention checks in the survey. We utilized a within-person design, with each participant invited to complete up to three randomly assigned vignettes. To minimize the potential influence of spill-over effects, or response tendencies, from one vignette to another vignette, restrictions were in place to ensure participants did not receive combinations of vignettes that were confusing or contradictory (e.g., participants would not have received a combination where an AI decision maker made both a positive and a negative decision within the same HRM function). Also, the order of the vignettes that participants read was randomized, which minimized the potential for order effects. Participants were also instructed to read each vignette independently of the others. After reading each vignette, participants were invited to complete survey measures (detailed below).

3.3 Sample

We recruited 638 North American participants to take part in the 20 minute survey in exchange for US$5.00. Upon reviewing the attention checks embedded in the survey, we removed 192 participants who failed to correctly answer one or more of our 24 attention checks (our most stringent cleaning process), resulting in a final sample of 446 individuals who completed the survey. Participants were, on average, 38.68 years old (SD = 12.99), 63.40% were female, 36.20% were male, and 0.40% were non-binary. Most of our sample had University degrees (33.1% with undergraduate qualifications and 26.4% with graduate or postgraduate degrees) and 81.1% identified as Caucasian (with 8.7% Black or African American as the next most common). In terms of marital status, 48.40% were married, 10.20% were in a de facto relationship, and 41.30% were single. The average work experience of respondents was 18.97 years (SD = 12.85), while their average company tenure was 7.25 years (SD = 7.60). In terms of employment, 73.60% of our sample worked full-time, 94.50% held a permanent position, and 46.90% were in a management position. Our respondents came from a wide range of sectors (top five listed here): health services (11.40%); professional services (10.60%); education (9.40%); food, drink, and tobacco (9.40%); and construction (9.10%).

3.4 Measures

We utilized the following measures for our variables of interest (collected after each vignette) and control variables (collected once at survey end), with all demonstrating good reliability. To inform our overall research question and hypotheses, participants were also invited to provide open-ended qualitative responses following the survey questions regarding interactional justice (e.g., "can you provide further details as to why you thought you were or were not treated with dignity and respect in this scenario?").

Interactional justice was measured by Colquitt's (2001) four item scale (1 = to a small extent and 5 = to a large extent). An example item is: "Has < the manager/the AI algorithm > treated you with dignity?" ($\alpha = 0.87$).

Decision-maker role appropriateness was measured by a single item constructed by the authors (1 = very inappropriate and 7 = very appropriate). The item was: "In this scenario, how appropriate is it to have a manager/AI algorithm make this decision?".

Trust was measured by Körber's (2018) two item scale (1 = strongly disagree and 5 = strongly agree). An example item is: "I trust the manager/AI algorithm" (α = 0.87).

Dehumanization was measured by Bastian and Haslam's (2011) five item scale (1 = strongly disagree and 5 = strongly agree) focused on feelings of mechanistic dehumanization. An example item is: "The < manager/AI algorithm > is treating me as if I were an object" (α = 0.88).

Control variables that included demographic variables (age, gender, education, marital status, cultural group) and work experience (years of experience, tenure at current company, current industry, type of employment) were collected. Other control variables collected were: propensity to trust in technology (Mcknight et al.,'s, 2011 seven item measure; α = 0.85); propensity to trust automated systems (Körber's, 2018 three item measure; α = 0.81); propensity to trust people (Mayer & Davis', 1999 eight item measure; α = 0.73); and algorithm aversion (three items from Melick's, 2020 measure; α = 0.75).

3.5 Analytical Strategies

The data were entered into SPSS Statistics 25.0 for MANOVA analysis and SPSS Modeler 18.2.1 (IBM_Corp, 2021) to generate machine learning models. The qualitative open-ended responses were entered into NVivo (version 12) for thematic coding and analysis.

As each individual was invited to complete up to three vignettes (within-person design), the unit of analysis is the number of completed vignettes rather than the number of respondents, thus resulting in a total of 1059 observations. Of those 1059 observations, only 759 had complete data. Given the nature of each analytical technique, the MANOVA analysis was based on the observations without any missing data ($n = 759$) and the machine learning analysis was based on all observations including those with missing data ($n = 1059$). To support the appropriateness of our quantitative analytical approaches given our within-person design, we calculated that the largest percentage of the variance in our variables of interest was attributable to between-person differences (ICC values were all below 0.05), thus indicating that a within-person analytical approach or multi-level analytical approach were not warranted (Maas & Hox, 2005). As the within-person variance was smaller than 5%, this further indicates that any variance due to participants completing multiple vignettes is close to zero, and this offers evidence that there was a negligible effect (that can therefore be ignored) from the same person reading multiple vignettes. The details of each analytical approach are now outlined.

3.5.1 MANOVA analyses

We conducted a MANOVA with our four groups (see Table 1) as the independent variable and our measures of interactional justice, role appropriateness, trust, and dehumanization as our dependent variables. We included the above-mentioned control variables in our analysis. Note that we conducted Bonferroni corrected ($p < 0.008$ based on six pairwise comparisons) post-hoc contrast analyses.

3.5.2 Machine learning analyses

To take a bottom-up approach to modelling (i.e., data driven using machine learning) to identify salient features in the data to predict perceived interactional justice, we used the C5.0 classification modelling methods in IBM SPSS Modeler. C5.0 is widely used for classification problems due to its efficiency and ease of interpretation (Han & Kamber, 2011). The C5.0 algorithm is based on the notions of information gain and entropy to build a *decision tree* or a *rule set* by iteratively splitting the dataset on the field that provides the maximum *information gain*. Pruning is used to remove lower-level splits that do not contribute to the model. As a classification algorithm, input fields are used to predict the value of a target field. The predictors, identified from among the input fields, are ranked by their importance for the creation of the decision tree. In the full analyses, we have used each of the four factors as target variables (interactional justice, role appropriateness, trust, and dehumanization) to learn what input variables are most important in predicting the target. Because the factors are averages of multiple Likert scale questions, it was necessary to convert these variables from continuous to categorical values. For example, for factors with five point Likert scales, three categories were created using the following splits: low/disagree = 1–2.5; medium/neutral = > 2.5 & < 3.5; high/agree = 3.5 and above. We repeated these analyses separately for each of the four groups, and all groups combined, thus creating 20 models in total. To avoid biasing our data we included all variables as input variables except where they were clearly irrelevant or already captured in another variable. We conducted tenfold-cross validation for each model.

3.5.3 Qualitative data: Open-ended survey response analyses

The qualitative data obtained from the free text responses was thematically analysed to identify themes within the data (Braun & Clarke, 2006). We adopted a bottom-up "inductive analysis" to allow themes to emerge organically from the data (Braun & Clarke, 2006, p. 83; Pratt, 2009). Themes were identified at a "latent or interpretative level" (Braun & Clarke, 2006, p. 84) by coding whole passages with mentioned themes. When participants mentioned multiple or contradictory themes within a single passage, we coded the passage with all relevant themes. We used "investigator [or researcher] triangulation" to ensure that different perspectives

Table 2 Means, Standard Deviations, and Correlations Among the Focal Variables

	M	*SD*	1	2	3	4	5	6	7	8
1. AI- Group (AI – negative)	-	-	-							
2. AI + Group (AI – positive)	-	-	-	-						
3. H- Group (Human – negative)	-	-	-	-	-					
4. H + Group (Human – positive)	-	-	-	-	-	-				
5. Interactional justice	3.31	1.11	-0.19***	-0.01	-0.13***	0.31***	-			
6. Role appropriateness	4.51	1.87	-0.29***	-0.03	0.01	0.31***	0.60***	-		
7. Trust	3.21	1.10	-0.25***	0.06	-0.12***	0.31***	0.68***	0.69***	-	
8. Dehumanization	3.14	0.98	0.27***	0.16***	-0.05	-0.37***	-0.47***	-0.47***	-0.49***	-

Notes. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. AI- Group contains 236 useful responses; AI + Group contains 276 useful responses; H- Group contains 272 useful responses; and H + Group contains 275 useful responses.

Table 3 Overview of MANOVA Results for Interactional Justice

	AI-	AI +	H-	H +
AI-	**2.92 (AI-)**	2.92 (AI-) < 3.29 (AI +)	2.92 (AI-) = 3.06 (H-)	2.92 (AI-) < 3.88 (H +)
AI +	3.29(AI +) > 2.92 (AI-)	**3.29 (AI +)**	3.29 (AI +) > 3.06 (H-)	3.29 (AI +) < 3.88 (H +)
H-	3.06 (H-) = 2.92 (AI-)	3.06 (H-) < 3.29 (AI +)	**3.06 (H-)**	3.06 (H-) < 3.88 (H +)
H +	3.88 (H +) > 2.92 (AI-)	3.88 (H +) > 3.29 (AI +)	3.88 (H +) > 3.06 (H-)	**3.88 (H +)**

Notes for Tables 3, 4, 5, and 6: Presented values are mean values on the outcome variable under investigation. The symbol ">" refers to significantly higher ($p < 0.001$) mean scores on the outcome variable; the symbol "<" refers to significantly lower ($p < 0.001$) mean scores on the outcome variable; and the symbol "=" refers to no significant difference in mean scores on the outcome variable.

Table 4 Overview of MANOVA Results for Role Appropriateness

	AI-	AI +	H-	H +
AI-	**3.54 (AI-)**	3.54 (AI-) < 4.41 (AI +)	3.54 (AI-) < 4.55 (H-)	3.54 (AI-) < 5.49 (H +)
AI +	4.41 (AI +) > 3.54 (AI-)	**4.41 (AI +)**	4.41 (AI +) = 4.55 (H-)	4.41 (AI +) < 5.49 (H +)
H-	4.55 (H-) > 3.54 (AI-)	4.55 (H-) = 4.41 (AI +)	**4.55 (H-)**	4.55 (H-) < 5.49 (H +)
H +	5.49 (H +) > 3.54 (AI-)	5.49 (H +) > 4.41 (AI +)	5.49 (H +) > 4.55 (H-)	**5.49 (H +)**

For details, see Notes under Table 3.

informed the thematic coding and to achieve inter-coder consistency (Carter et al., 2014). Our process involved two of the researchers independently coding the data from two vignettes and then developing a combined coding scheme that captured the range of themes in the data. One of the researchers then coded the entire dataset with the combined coding scheme.

4 Results

4.1 Descriptive Statistics

Table 2 provides an overview of the means, standard deviations, and correlations among the study variables (interactional justice, role appropriateness, trust, and dehumanization). In line with our hypotheses, these correlations are presented in accordance with the overview presented in Table 1: AI- Group (AI is decision maker and decides negative); AI + Group (AI is decision maker and decides positive); H- Group (human is decision maker and decides negative) and H + (human is decision maker and decides positive). Table 2 also overviews the number of completed vignettes per group.

4.2 MANOVA Results

All MANOVA results are presented below in the following order: we first focus on differences in *who* the decision maker was (AI versus human), and then focus on differences in the *valence* (negative versus positive) of the decision. The results for each focal variable are presented in Tables 3, 4, 5 and 6.

Table 5 Overview of MANOVA Results for Trust

	AI -	AI +	H-	H +
AI -	**2.72 (AI -)**	2.72 (AI -) < 3.31 (AI +)	2.72 (AI -) < 2.98 (H-)	2.72 (AI -) < 3.78 (H +)
AI +	3.31 (AI +) > 2.72 (AI-)	**3.31 (AI +)**	3.31 (AI +) > 2.98 (H-)	3.31 (AI +) < 3.78 (H +)
H -	2.98 (H-) > 2.72 (AI-)	2.98 (H-) < 3.31 (AI +)	**2.98 (H-)**	2.98 (H-) < 3.78 (H +)
H +	3.78 (H +) > 2.72 (AI-)	3.78 (H +) > 3.31 (AI +)	3.78 (H +) > 2.98 (H-)	**3.78 (H +)**

For details, see Notes under Table 3.

Table 6 Overview of MANOVA Results for Dehumanization

	AI-	AI +	H-	H +
AI-	**3.62 (AI-)**	3.62 (AI-) > 3.42 (AI +)	3.62 (AI-) > 3.05 (H-)	3.62 (AI-) > 2.53 (H +)
AI +	3.42 (AI +) < 3.62 (AI-)	**3.42 (AI +)**	3.42 (AI +) > 3.05 (H-)	3.42 (AI +) > 2.53 (H +)
H-	3.05 (H-) < 3.62 (AI-)	3.05 (H-) < 3.42 (AI +)	**3.05 (H-)**	3.05 (H-) > 2.53 (H +)
H +	2.53 (H+) < 3.62 (AI-)	2.53 (H +) < 3.42 (AI +)	2.53 (H +) < 3.05 (H-)	**2.53 (H +)**

For details, see Notes under Table 3.

4.2.1 Interactional justice results

As shown in Table 3, results of our MANOVA indicated a significant difference between our four groups in interactional justice perceptions [$F(3, 182) = 21.68$, $p < 0.001$, $\eta 2 = 0.18$]. In terms of who the decision maker was (H1a), we found that: (1) when a decision is positive, interactional justice perceptions are significantly higher if the decision is made by a human rather than an AI; (2) when a decision is negative, there is no significant difference in interactional justice perceptions when the decision is made by a human or an AI; (3) interactional justice perceptions are higher when a human makes a positive decision compared to an AI making a negative decision; and (4) interactional justice is higher when an AI makes a positive decision compared to a human making a negative decision. In terms of the valence of the decision (H2a), we found that: (1) irrespective of who makes the decision (AI or human), interactional justice perceptions are higher if the decision is positive rather than negative; and (2) thus, for our exploratory case, interactional justice perceptions are higher when an AI makes a positive decision compared to a human making a negative decision.

4.2.2 Role appropriateness results

As shown in Table 4, results of our MANOVA indicated a significant difference between our four groups in role appropriateness perceptions [$F(3, 182) = 76.39$, $p < 0.001$, $\eta 2 = 0.14$]. In terms of who the decision maker was (H1b), we found that: (1) for both positive and negative decisions, role appropriateness scores are higher when the same decision is made by a human rather than an AI; (2) role appropriateness scores are higher when a human makes a positive decision compared to an AI making a negative decision; and (3) there are no differences in role appropriateness scores when a human makes a negative decision compared to an AI making a positive decision. In terms of the valence of the decision (H2b), we found that: (1) irrespective of who makes the decision (AI or human), role appropriateness scores are higher if the decision is positive rather than negative; (2) role appropriateness scores are higher when a human makes a positive decision compared to an AI making a negative decision; but (3) in our exploratory case, there is no significant difference in role appropriateness scores when a human makes a negative decision compared to an AI making a positive decision.

4.2.3 Trust results

As shown in Table 5, results of our MANOVA indicated a significant difference between our four groups in trust perceptions [$F(3, 182) = 21.39$, $p < 0.001$, $\eta 2 = 0.13$]. In terms of who the decision maker was (H1c), we found that: (1) for both positive and negative decisions, trust scores are higher when the same decision is made by a human rather than an AI; (2) trust scores are higher when a human makes a positive decision compared to an AI making a negative decision; and (3) trust scores are lower when a human makes a negative decision compared to an AI making a positive decision. In terms of the valence of the decision (H2c), we found that in all cases positive decisions have higher trust scores than negative decisions, including in our exploratory case.

4.2.4 Dehumanization results

As shown in Table 6, results of our MANOVA indicated a significant difference between our four groups in dehumanization perceptions [$F(3, 182) = 24.06$, $p < 0.001$, $\eta 2 = 0.18$]. For the dehumanization variable, lower scores mean lower feelings of dehumanization. In terms of who the decision maker was (H1d), we found that: (1) dehumanization scores are lower when the same decision is made by a human rather than an AI, both for positive and negative decisions; (2) dehumanization scores are lower when a human makes a positive decision compared to an AI making a negative decision; and (3) dehumanization scores are lower when a human makes a negative decision compared to an AI making a positive decision. In terms of the valence of the decision (H2d), we found that irrespective of who makes the decision (AI or human), dehumanization scores are lower if the decision is positive rather than negative, excluding in our exploratory case.

4.3 Machine Learning Results

To inform our main research question, we only present here the results for the model (decision tree) produced that used as input from the dataset all four groups and the target variable of interactional justice. Our goal was to learn which variables predict the three interactional justice classes: *LOW* (responses from 1 to 2.5); *MEDIUM* (responses over 2.5 and less than 3.5); and *HIGH* (responses 3.5 and above). This resulted in a tree depth of 8, accuracy mean of 61.8 and Standard Error of 1.1. We note that differences were found when models were created for the four groups separately and when different target variables were used, but the results consistently indicated that the same variables were important in predicting the outcome. Table 7 shows the predictor importance produced by SPSS using the C5.0 classification algorithm with interactional justice as the target variable with three classes (*LOW, MEDIUM* and *HIGH* per the descriptions above). These 10 variables can predict 95.16% of the data. This shows us that trust is the most salient variable, followed by role appropriateness and then dehumanization, which can predict 32.23%, 29.23% and 9% of the observations, respectively. Consistent with Table 7, we can see from Fig. 1 that trust, role appropriateness, and dehumanization are the key features in predicting *LOW*, *MEDIUM*, or *HIGH* interactional justice; modes 1, 2 and 3, respectively. We can see (in line 2 in Fig. 1) that very low trust (less than 1.75 on a 5 point Likert scale) is sufficient to predict *LOW* perceived interactional justice for 100 observations with 86% accuracy. When trust is greater than 1.75, interactional justice will be perceived as *MEDIUM* if the person does not feel dehumanized (< = 3.5 on a five point Likert scale), with this being the case for 75 observations, or *LOW* when they do feel dehumanized (> 3.5) as found in 86 observations. In contrast, we see that 353 observations with 78.8% accuracy can be predicted to perceive interactional justice as *HIGH* when trust is medium or above (> 2.75) and role appropriateness is high (> 5.5 on a seven point Likert scale).

Table 7 Predictor Importance for Interactional Justice as Target Variable

Input Variables	Importance
Trust (in decision maker)	0.3223
Role appropriateness	0.2923
Dehumanization	0.0911
Trust in humans (generally)	0.0588
Age	0.0537
Vignette	0.0478
Overall tenure	0.0279
Cultural group	0.0276
Trust in technology (generally)	0.018
Algorithm aversion	0.0121
Total predictiveness	0.9516

Notes. To gain an understanding of what variables are relevant for predicting *LOW*, *MEDIUM*, or *HIGH* interactional justice, we need to look at the decision tree. While the full tree has a depth of up to eight, for clarity and simplicity we present the tree showing a depth of three in Fig. 1.

The plus sign in the decision tree in Fig. 1 indicates that the branch can be expanded further, which is necessary to identify the variables that cause further splits, the class (which might be different to the higher-level node), number of observations, and accuracy. In expanding the decision tree, we found that if trust is medium (between > 2.75 and < = 3.25) and role appropriateness is low (< = 2.5) then interactional justice perceptions are *LOW* (21 responses with at least 71.4% accuracy). In general, the expanded decision tree reveals that trust, dehumanization, or role appropriateness are enough to classify over half of the responses (635 out of 1059 observations), but that other factors will depend on the specific vignettes and participants' individual factors such as their culture, age, or attitude to technology. For example, when trust is not low (> 1.75) but feelings of being dehumanized are high (> 3.5), self-identified 'white' participants who found role appropriateness to be low to medium (< = 4.5) will consider interactional justice to be *LOW* (66 observations with 66.7% accuracy), but if role appropriateness is high (> 4.5) then interactional justice is perceived by them as *HIGH* (5 observations with 60% accuracy). Self-identified 'white' participants will consider interactional justice to be *MEDIUM* if trust is not low (> 1.75) and feelings of dehumanization are not high (< = 3.5) (17 observations with 76.5% accuracy).

Fig. 1 Decision Tree for Interactional Justice: Branch Depth of 3.
Notes. Mode 1 = *LOW*; Mode 2 = *MEDIUM*; Mode 3 = *HIGH*. Number in brackets (number of observations, accuracy)

In total there are 78 leaf nodes in the decision tree. Often the number of responses covered by specific combinations of conditions is small, usually with less than 10 responses, and thus it is likely that another set of data would produce different results for these cases. Looking at the leaves with over 10 observations and generalizing the rules, we can conclude that interactional justice is:

- *LOW* if: (1) trust is low (< = 1.750); OR (2) trust is not low (> 1.750) but people feel dehumanized (> 3.500); OR (3) trust is medium (> 2.750 and < = 3.25) but appropriateness of the decision maker is low (< = 2.500);
- *MEDIUM* if: (1) trust is not low (> 1.750) and people feel moderately dehumanized (< = 3.500); OR (2) trust is medium or above (> 2.750), the decision maker is viewed as moderately appropriate (> 2.500 and < = 5.500), and people feel moderately to highly dehumanized (> 2.500);
- *HIGH* if: (1) trust is medium or above (> 2.750) and role appropriateness is high (> 5.500); OR (2) trust is high (> 3.250) even though role appropriateness is low (< = 2.500); OR (3) trust is medium or above (> 2.750), role appropriateness is medium (> 2.500 and < = 5.500), but feelings of dehumanization are low (< = 2.500).

4.4 Qualitative Data: Open-Ended Survey Responses

The results are presented across two tables reflecting our coding structure. All relevant data were coded under one or more minor themes (see Table 8). We then grouped these minor themes under the three major themes of 'negative', 'positive', or 'neutral' for justice. Descriptions and illustrative quotes for our minor themes are presented in Table 8.

We focus our discussion on Table 9, which reports the frequency with which each theme was used. The bold numbers represent the cumulative percentages for the three major themes (positive, negative, or neutral for justice) and the remaining numbers indicate how often, as a percentage, each minor theme was used across each of the four groups (AI-, AI+, H-, H+).

The *human positive group* (H + Group) had by far the largest percentage of themes that were positive for justice (87.2%). The most common positive themes were that: (1) the decision was based on relevant data (32.8%; e.g., the decision was fair because it "used performance data" [H+] and was "data driven" [H+]); (2) a fair outcome had been achieved (24.2%; e.g., "I deserved it" [H+]); and (3) the decision maker was respectful (16.1%). This group had, by far, the lowest percentage of themes that were negative (7.9%) and neutral (4.8%) for justice.

The *human negative group* (H- Group) had more negative (48.5%) than positive (40%) themes for justice. This group had far fewer positive and far more negative themes than the human positive group, but far more positive and far fewer negative themes than the AI negative group. It also had less positive and more negative themes than the AI positive group, although the gap here was smaller. The most common positive theme was again that the decision had been made based on relevant data (17.1%). This suggests that it is important, in terms of justice, to be clear what data are used to base decisions on and the comprehensiveness and relevance of that data. The most common negative theme was that a bad or unfair outcome had occurred (18.8%; e.g., it "was unfair because I deserved the job" [H-]).

The AI positive group (AI + Group) had more positive (52.8%) than negative (33.1%) justice themes, although this gap was much smaller than in the human positive group. This suggests that even with a positive decision, there were a number of negative justice themes raised in the AI positive group. As with all other groups, the most common positive theme was that the decisions were being driven by appropriate and relevant data (19.2%; e.g., "it was determined based purely on facts" [AI+]). This suggests that people are willing to accept a positive AI decision as fair if they view the data the decision was based upon as relevant and reliable. The two most common negative themes were that it was not appropriate for an AI to make that sort of decision (9.4%; e.g., that is not a "decision that should be taken by the algorithm" [AI+]) and that the decision was based on irrelevant or flawed data (8.9%; e.g., "it's a slightly cold way to analyse

Table 8 Themes Identified in the Qualitative Data

Theme Description	Illustrative quotes
Negative for justice	
Bad or unfair outcome; didn't deserve it	"*You deserve it and it was not given to you*" [H-] "*It is just not fair*" [AI+] "*Because it didn't work out in my favor*" [H-]
Decision maker is not trustworthy	"*I don't trust that this algorithm is trustworth*y" [AI+]
Decision maker not competent or able to make the decision	"*I don't see how an algorithm could makes these decisions*" [AI-] "*How is an AI algorithm going to detect personality and likeability*?" [AI-]
Decision maker lacks emotional intelligence or emotions	"*Unfair because it does not take feelings and hard work into consideration*" [AI-] "*AI is incapable of such emotions*" [AI+] "*Computers have no emotion intelligence*" [AI-]
Decision maker was (or could be) biased	"*Maybe my manager has other agenda that wasn't made known to me*" [H-] "*Algorithms can have bias*" [AI+] "*The manager could have favoritism and the system could be rigged*" [H-]
Decision maker was disrespectful; dehumanizing; no dignity	"*Wasn't treated as an individual with feelings*" [AI+] "*I am just a number not a human to it*" [AI-] "*I feel as if I were treated like an object*" [H-] "*I wasn't treated with dignity because my performance at the company wasn't appreciated*" [H-]
Decision based on wrong or irrelevant data or missed relevant data; relevant information is not quantifiable	"*It does not take into account your actual personality, work ethic, drive & ambition. Things that you can't quantify. These things only a human could pick up on*" [AI-] "*Because the algorithm does not know me as human only statistics provided by computer*" [AI-] "*Good employees can be more than just good data*" [H-]
Human needs are not being met by that decision	"*Because I really need the job, but the algorithm rule is stopping me*" [AI-] "*Because some other staffs will lose their means of livelihood*" [H+] "*It didn't respect my needs*" [AI-]
Lack of explanation for the decision	"*I want a reason why I am not good enough for the promotion*" [H-] "*I deserve an explanation*" [H-] "*They did not explain why you should not be offered the job*" [AI-]
Lack of recourse to query the decision	"*No chance to even have the schedule be reconsidered*" [AI-] "*Does not seem I was given the courtesy to argue my case*" [H-]
Not appropriate for the decision maker to make that decision	"*I don't think [this] is a decision that should be taken by the algorithm*" [AI+] "*A real person [not an AI] should make that decision*" [AI-]
Neutral for justice	
Lack information to answer	"*Not enough info to determine*" [H+] "*There are not enough facts to make an opinion*" [H+]
Neither respected nor disrespected; neither fair nor unfair	"*It's [AI] not a real person. Impossible [for it] to be respectful or disrespectful*" [AI+] "*He was neither fair or not fair*" [H+] "*An algorithm doesn't have feelings, so it can't treat you with respect or dignity, but it also can't treat you with disrespect*" [AI+]
Positive for justice	
Appropriate for the decision maker to make the decision	**"***It is better [to have an AI deciding] than a human deciding*" [AI+] "*The manage[r] has the human ability to make such a determination*" [H+] "*It's a real person making the decision, not a computer*" [H+]
Decision based on relevant data; data driven decision; fair information-based criteria used	**"***This is a perfect way of analyzing the current skills and performance records to give a precise answer to whom is qualified*" [AI+] "*I was evaluated based on facts exactly the same way the other candidates were evaluated*" [AI-] "*The manager used data to justify the decision*" [H+]

Table 8 (continued)

Theme Description	Illustrative quotes
Decision maker is not biased	**"***The algorithm is looking at everyone's skills. There is no bias when it is an algorithm*" [AI+] "*Was an unbiased, fact-based decision*" [AI+] "*The manager seemed to have been impartial using just the data to decide*" [H+]
Decision maker is respectful; treats with dignity; treats equally; there was no disrespect	**"***I was treated with respect, I feel the data was looked at and I was chosen*" [AI+] "*The manger treated them with the most dignity*" [H+] "*I'm treated with dignity because my assessment is based on capability and experience*" [H+] **"***The algorithm is respectful because it takes no personal traits into account*" [AI-] "*I felt as though my manager was equal to me and they knew they were equal to me*" [H+]
Decision maker is trustworthy	**"***I trust my manager made a fair decision based on my past skills and job performance*" [H+] "*I trust the algorithm*" [AI+]
Good or fair outcome; deserved it	**"***I got to take the training, therefore no problems*" [AI+] "*I deserved it*" [H+] "*It seemed fair*" [AI+]
Human needs are being met by that decision	**"***It chose participants that are really in need of the training*" [AI+] "*It was fair because the worker needed money so it helps in reducing cost*" [AI+] "*They considered the needs of the employees*" [H+]

Notes. The group the quote comes from is indicated as AI+, AI-, H+ and H-.

Table 9 Frequency of Theme Usage as a Percentage

Theme	AI- Group %	AI+ Group %	H- Group %	H+ Group %
Negative for justice	**65**	**33.1**	**48.5**	**7.9**
Bad or unfair outcome	9.9	2.6	18.8	2.4
Decision maker is not trustworthy	0.5	0.2	0	0
Decision maker not competent or able to make decision	3.8	1.4	0	0
Decision maker lacks emotional intelligence or emotions	4.6	4.2	0	0.4
Decision maker was (or could be) biased	0.3	0.9	3.2	2
Decision maker was disrespectful	7.9	3.8	6.1	1.3
Decision based on wrong or irrelevant data or missed relevant data	17.8	8.9	6.8	1.5
Human needs are not being met by that decision	3	1.2	5.1	0.2
Lack of explanation for the decision	2.5	0.5	6.3	0
Lack of recourse to query the decision	0.8	0	1	0
Not appropriate for the decision maker to make the decision	14	9.4	1.2	0
Neutral for justice	**14.5**	**14.1**	**11.5**	**4.8**
Lack information to answer	1.3	0.7	8	3.1
Neither respected nor disrespected; neither fair nor unfair	13.2	13.4	3.4	1.8
Positive for justice	**20.6**	**52.8**	**40**	**87.2**
Appropriate for the decision maker to make the decision	0.3	0.9	6.8	5.7
Decision based on relevant data; it is a data driven decision	9.1	19.2	17.1	32.8
Decision maker is not biased	2.5	4.7	0.7	2
Decision maker is respectful	3.8	7.5	8.3	16.1
Decision maker is trustworthy	0	1.2	0.2	0.4
Good or fair outcome	4.3	16	6.8	24.2
Human needs are being met by that decision	0.5	3.3	0	5.9

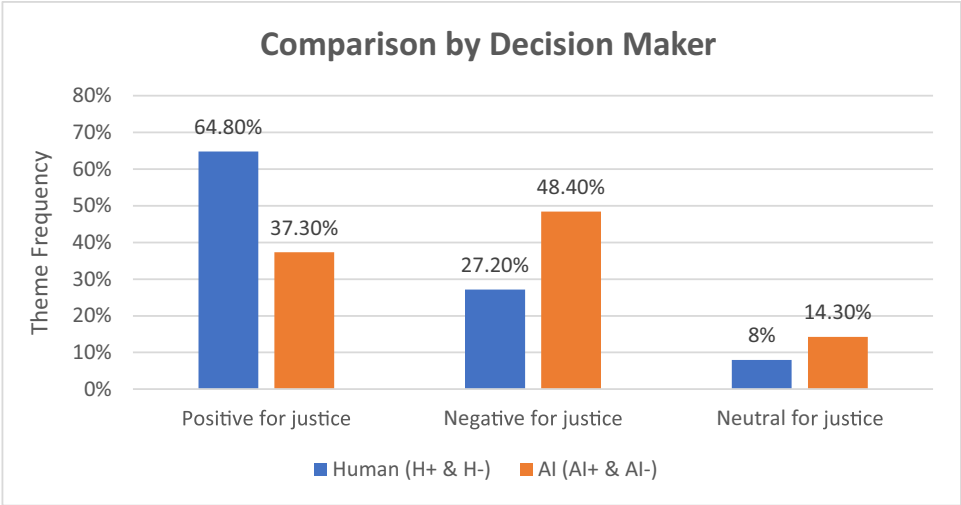

Fig. 2 Theme Frequency as a Percentage Comparison by Decision Maker

workers' performances because there could be other factors at play" [AI+]).

The AI negative group (AI- Group) had the highest percentage of negative justice themes and the lowest percentage of positive justice themes of any group. As with the AI positive group, the two most common negative themes were the decision being based on irrelevant or flawed data (17.8%) and the AI not being an appropriate decision maker (14%; e.g., "a real person should make that decision" [AI-]). For the former theme, this was commonly expressed in terms of some human features being unquantifiable or not reducible to maths (e.g., "it does not take into account your actual personality, work ethic, drive & ambition … things that you can't quantify" [AI-]). For the positive themes, the most common was that the decision was data driven (9.1%).

Several other insights are generated from this data. The theme of being neither respected nor disrespected was very low for both human cases (3.4% for H- and 1.8% for H+), compared to the AI cases (13.2% for AI- and 13.4% for AI+). For the two AI groups, this was typically expressed as it being "impossible [for an AI] to be respectful or disrespectful" [AI+] or that an AI is "neither fair [n]or unfair" as it is simply "working according to its programming" [AI-]. Although there are many concerns raised about AI bias in ethical AI literature (e.g., Kellogg et al., 2020), our data suggests that this has the potential to gloss over bigger concerns about human bias that were raised by participants. We found that participants were, comparatively, more likely to praise AI's perceived lack of bias (3.7% for AI+& AI- vs. 1.4% for H+& H-) in its decision making (e.g., "there is no bias when it is an algorithm" [AI+]), and more likely to raise concerns (2.5% for H+& H- vs. 0.6% for AI+& AI-) about human bias (e.g., "people will sometimes play favorites regardless of the data" [H+]). A lack of emotion or emotional intelligence emerged as a theme for both AI groups (e.g., "the AI had no human feelings" [AI+]), but this was (understandably) barely mentioned in both human groups. The fact that a human was an appropriate decision maker (6.3% for H+& H-) was an identified theme (e.g., "the manager is another human with emotions and can fairly evaluate the situation" [H+]), whereas an AI was rarely mentioned as an appropriate decision maker (0.6% for AI+& AI-). Instead, the AI was mentioned more (11.6% for AI+& AI-) as an inappropriate decision maker (e.g., "computers shouldn't be making human decisions" [AI-]), whereas rarely were humans mentioned as inappropriate decision makers (0.6% for H+and H-).

In general, we see that people respond positively, in terms of justice themes, to human over AI decisions and to positive over negative decisions, as shown in Figs. 2 and 3, where we separate the results by decision maker (Fig. 2; H+& H- vs. AI+& AI-) and by decision valence (Fig. 3; H+& AI+vs. H- & AI-). In terms of our exploratory case, comparing AI positive and human negative groups shows that the AI positive group had higher positive justice themes (52.8% vs. 40%) and lower negative themes (33.1% vs. 48.5%) compared to the human negative group. This suggests that, in terms of the frequency of themes, whether the decision was positive was more important than who constituted the decision maker. Figure 2 also shows that for the two human groups combined, justice perceptions are overall quite positive (64.8% positive vs. 27.2% negative), whereas for the two AI groups combined, justice perceptions were more negative than positive (37.3% positive vs. 48.4% negative). Similarly, for the two positive groups combined (Fig. 3), the justice perceptions were more positive (70.6% positive vs. 20.1% negative), whereas for the two negative groups combined the justice perceptions were more negative than positive (30.5% positive vs 56.6% negative).

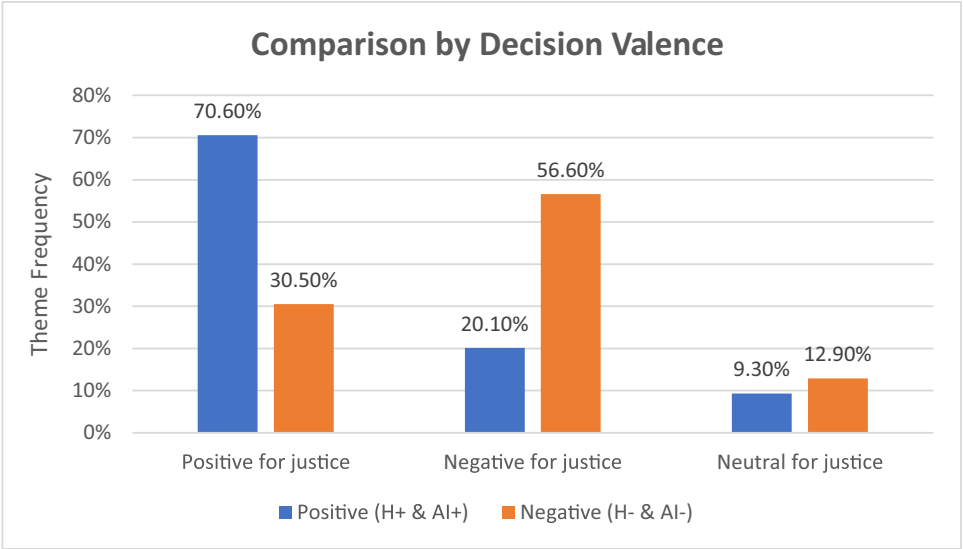

Fig. 3 Theme Frequency Percentage Comparison by Decision Valence

Although the decision-making data presented in the vignettes were identical across all groups, there were clear differences in how the use of that data in the decision-making process was assessed. For the two positive valence groups (AI + & H +), basing decisions on this data was mostly seen as a positive for justice as it meant the decisions were "data driven" (26.3% for AI + & H + vs. 13.2% for AI- & H-). In contrast, for the two negative valence groups (AI- & H-), basing decisions on this data was seen as using irrelevant or flawed data or as missing important relevant data (12.2% for AI- & H- vs. 5.1% for AI + & H +). This suggests that when a decision is positive, people tend to rationalize the decision as being based on appropriate data that vindicates the positive decision, and when the decision is negative people are more likely to rationalize the decision as being based on inappropriate data or due to missing relevant data.

5 Discussion, Limitations, and Future Research Directions

As AI undertakes more decision making in organizations, it is necessary to understand people's responses to this shift. This is particularly important where AI assumes decision making in domains that can have significant impacts upon human well-being, such as HRM-related decisions. Our work provides insights into an important facet of people's experiences of AI decision making in a HRM context: their experiences of respectful and dignified treatment through interactional justice and related perceptions of decision-maker appropriateness, trust, and dehumanization.

5.1 Theoretical Implications

Our findings make four contributions to this field of study. First, outcome bias (where evaluations of a decision focus more on its outcomes than its processes) is a well-documented phenomenon (Lipshitz, 1989). In offering evidence of this phenomenon in the context of AI decision making we reveal how people construe the positive and negative implications of AI for their experiences of interactional justice. Our quantitative work shows that people are generally less trusting, more dehumanized, and experience less interactional justice when a decision is made by an AI (compared to a human), and they often view the AI as an inappropriate decision maker. Our qualitative work surfaces the specifics of these negative cognitions, with AI viewed as (1) an inappropriate decision maker, (2) basing its decision on wrong or irrelevant data, (3) being disrespectful (or unable to express respect or disrespect), (4) lacking emotional intelligence, and (5) not competent or able to make the decision. However, our quantitative results also show that when AI makes a positive decision (when compared to either an AI or a human making a negative decision), people felt more trusting and had higher interactional justice perceptions. In this aspect, our qualitative work helps to uncover why this occurs as people view AI, in this case, as (1) making an appropriately data-driven decision, (2) being respectful, and (3) being unbiased. This suggests the use of the "machine heuristic", where automated systems are viewed as universally unbiased, objective, and consistent (Araujo et al, 2020, p. 612).

Examining this outcome bias highlights the tensions in people's understanding of AI and this contributes to what Logg et al., (2019, p. 100) term the "*theory of machine*", or our knowledge of "people's lay perceptions of how

algorithmic and human judgment differ in their input, process, and output". Our participants could identify both the benefits (via activation of the "machine heuristic") and limitations (a converse "anti-machine heuristic") of AI, and they tended to focus more on the former when AI affords them a positive decision, while they tended to focus more on the latter when AI affords them a negative decision. These findings demonstrate a complexity beyond a positive–negative view of AI and help to expose when and why people activate both positive views (e.g., it is unbiased) and negative views (e.g., it cannot capture all relevant information) of its capacities. However, as the literature suggests (Lipshitz, 1989), outcome bias can cloud one's judgments of the *process* taken toward a decision. This implies that individuals may not be properly evaluating how an AI reaches its decisions but are instead focusing on the outcome they receive from that decision. Interestingly, these findings also hold across the six different HRM functions we focused upon. That is, we found no significant differences across the six scenarios for all our outcome variables. While other work suggests that humans may view AI decision making differently depending on the task (e.g., human or mechanical tasks, see Lee, 2018), our results offer evidence that people's views of AI decision making remained consistent regardless of the type of HRM decision at stake. These results suggest that individuals may not have a particularly nuanced understanding of the workings of AI, which is problematic as its deployment for decision making is accelerating across industries (Kellogg et al., 2020).

The machine learning results extend our knowledge of how individual characteristics will also play a role in how people construe AI decision making. For example, cultural background, among other factors, was shown to play a role in individuals' justice perceptions. This supports other work such as Gupta et al. (2021) who, using established cultural dimensions, provide evidence that a manager's cultural identity may impact on the extent to which they accept or question an AI-based decision. They suggest that a manager who is individualistic with low masculinity and weak uncertainty avoidance is more likely to accept a recommendation without question, compared to a manager who has a collectivist orientation with high masculinity and strong uncertainty avoidance. These findings highlight the need for a better understanding of what individual characteristics influence attitudes toward AI decision making.

Our second set of contributions relate to bias in decision making. There is an increasing and justifiable focus in academic literature and the popular press on issues of AI bias. Our work offers novel insights into how people construe AI versus human bias in decision making and helps further advance our understanding of people's "theory of machine". That is, our qualitative analysis shows that, compared to AI decision making, participants raised more concerns about human biases impacting decisions (such as managers who "play favorites regardless of the data" [H+]), and often referenced a lack of AI bias in its decision making (e.g., it was "an unbiased, fact-based decision" [AI+]). While AI has been shown to replicate, at scale, many systemic and historic human biases that have unfairly marginalized specific groups, AI can also be free of other uniquely human biases and cognitive shortcuts like favoritism and post-hoc rationalizations. Research shows that these types of human biases can heighten worker stress and job dissatisfaction (Arasli & Tümer, 2008). Therefore, our work suggests that a predominant focus on AI bias may be marginalizing important conversations regarding the ongoing threat of human bias in decision making and the impact this continues to have on workers' experiences of interactional justice. However, activation of the "machine heuristic" that AI systems are largely unbiased represents a problematic aspect of people's "theory of machine". This heuristic may lead workers to uncritically accept AI decisions, particularly if they are positive as our quantitative work shows, without interrogating the process through which the decision was made, as outcome bias can cloud these assessments. This suggests an increasingly important role for discussions of the benefits of human-AI synergy or symbiosis (Jarrahi, 2018) that focus on humans and technology balancing each other's limitations and enhancing each other's strengths.

Third, an interesting insight emerged from the qualitative data regarding the capacity of AI to demonstrate respect. While our quantitative work showed that people could generate overall assessments of interactional justice in both AI and human decision-making contexts, our qualitative work added nuance to this finding. In the combined cases of AI as the decision maker (AI+ & AI-), several participants responded that the technology was unable to show respect or disrespect, which neutralized their subsequent views of interactional justice. Commentary such as the AI being "*neither fair [n]or unfair*" [AI-] and that it cannot treat "*someone [in a] disrespectful or respectful*" way [AI+] suggests that people may struggle to apply the language of fairness and respect to an AI. In essence, the technology "*does not treat you in any way, it [simply] analyses factual data*" [AI-]. These views appeared to be driven by the perceived non-human and unemotional nature of the AI. It may be that to fully assess interactional justice in AI decision making it is more important for people to assess the *process* through which a decision is reached. However, this may be hampered by our earlier evidence of an outcome bias that is shown to diminish the focus on decision processes.

Finally, our machine learning analysis demonstrates some of the complexity of the cognitions, and indeed the chain of potential relationships, that underpin individuals' experiences of AI decision making. While trust was shown to be a key predictor of interactional justice perceptions, the

saliency of this variable could vary based on the extent to which a participant perceived role appropriateness and dehumanization. These results suggest that individuals will assess multiple factors, apparently captured well by our focal variables, to construe overall interactional justice perceptions in our decision-making contexts.

5.2 Practical Implications

Our work also has several practical implications. Although human decision making was generally preferred, the positivity of the decision played a role. In particular, when an AI made a positive decision, people generally experienced higher interactional justice, higher trust, and lower dehumanization compared to a human or an AI making a negative decision. This suggests that organizations need to pay attention not just to *who* or *what* makes a decision, but also to the *positive* or *negative* outcome of a decision. We offer evidence that decisions with positive valence, regardless of the decision maker, will have fewer negative implications for workers' feelings of respectful treatment at work. Organizations also need to be aware that decisions with negative valence, especially negative decisions made by an AI, can generate feelings of disrespectful treatment, and they should attempt to limit or address those negative outcomes.

Our results also support other work (e.g., Araujo et al., 2020) showing an ongoing human hesitancy toward and distrust of the use of AI, particularly in sensitive decision-making areas such as HRM (Lee, 2018). This suggests that it may be beneficial for organizations to maintain humans in, or on, the loop for such decision making. These loop distinctions refer to humans continuing to have a meaningful role in either oversighting (on-the-loop) or actively working alongside (in-the-loop) AI in its decision making (Walsh et al., 2019), which could then support interactional justice perceptions.

Beyond the deployment of these technologies, there may also be an educative role for organizations in upskilling workers on the benefits and limitations of AI use. Managers, in particular, could benefit from training concerning the affordances of the technology, as well as an awareness of the dangers of perceiving AI as adding "extra workload" which has the potential to act as a "techno-stressor" which can influence AI justice perceptions and behavioral outcomes (Wang et al., 2021; cf. Yassaee and Mettler, 2019). There remains an ongoing lack of significant organizational investment towards transitioning and skilling workers in the use of AI (Halloran & Andrews, 2018), which may be exacerbating a poor understanding of how these technologies operate and how they can complement human skills. Such an educative role could help employees to assess AI decision making more accurately by moderating their use of a "machine heuristic" and support them to take a more active role in helping determine where AI can be best deployed in their organizations (Aizenberg and van den Hoven, 2020). In particular, more education of workers seems to be needed about the presence of algorithmic bias, the sort of data that AI systems use and don't use, the potential for "brittleness" in AI systems (McCarthy, 2007), and the dangers of uncritical reliance on AI. This could also be done in combination with efforts to tackle human bias in decision making, such as by blinding names when assessing CVs and efforts to reduce discrimination in the workplace. However, our machine learning results suggest that educational programs need to go beyond a one-size-fits-all approach as individual factors such as trust in technology and/or trust in humans, age, culture, and tenure will influence employees' beliefs and attitudes about the appropriate use of AI, how its use in decision making makes them feel, and their trust towards it. Finally, education should also both raise and address the ethical concerns of those impacted by AI decision-making technologies, such as assuring the presence of privacy safeguards (Kumar et al., 2021).

5.3 Limitations and Future Research Directions

Like all studies, our work comes with limitations. By its nature experimental work decontextualizes the phenomenon, meaning that we did not capture likely salient aspects of the workplace context such as leadership (e.g., the quality of leader-member relations), the presence of institutional decision-making policies (e.g., directives about who can make what decision based on what criteria), and broader work experiences (e.g., job satisfaction). Further, it may be more or less difficult for some participants, depending on their personal experience, to accurately predict their responses to unfamiliar situations. To address this, we picked common HRM situations which should be at least familiar to our working adult sample, and by randomly assigning participants to vignettes we minimized this issue. As discussed in our Methods section, we also had experts review our vignettes to maximize relevance and external validity. Again, driven by our method, our focus on manipulating decision maker (AI versus human) and decision valence (negative versus positive) dimensions sidelined salient aspects of AI that are known to impact humans' perceptions of it, such as the extent of its explainability (Hagras, 2018) and the transparency of its computations (Springer & Whittaker, 2020), although some of these issues did emerge in our qualitative analysis.

There remains much for future research to explore regarding humans' perceptions of AI decision making. While our study focused on unembodied (i.e., software-based) forms of AI, the increasing use of chatbots and avatars as embodied forms of the technology raise further questions regarding whether these types of AI increase people's perceptions

that the AI, as an anthropomorphized (i.e., an increasingly human-like) entity, can indeed treat them with respect or disrespect and whether computers "expressing feelings they can't have" is ethically appropriate (Porra et al., 2020). Araujo et al. (2020) note that when anthropomorphizing is lessened, an individual is more likely to view the technology as objective. It may be the case that when a technology is presented as more human-like, interactional justice is perceived differently than when it is presented as less human-like. Relatedly, pushes toward advancing "whitebox" AI that offers heightened explainability and transparency in the technology's computations (Hagras, 2018) may positively affect interactional justice perceptions. Bies and Moag's (1986) initial conceptualization of this form of justice focuses on both interpersonal and informational justice, with explainability likely important for enhancing the latter aspect. However, some research shows that "too much" explanation of AI's workings and decisions can be overwhelming for users (Walton, 2004) and potentially create more questions than it answers (Miller, 2019). Manipulating the amount and form of explanations provided for AI decision making would help unpack how this characteristic of the technology enhances or diminishes interactional justice perceptions. Future work could also explore the impact of various individual-level characteristics, such as those identified in our machine learning analysis (e.g., cultural background, age, attitude toward technology), on responses to different decision makers and decision valences. This would afford a more nuanced understanding of what factors influence attitudes towards AI decision making.

Finally, replicating our work in different contexts would be valuable. Our research focused on a specific decision-making context, but with governments moving to regulate the use of AI in high-risk domains (such as healthcare contexts, Balasubramanian, 2021), it is important to understand how people's perceptions of AI decision making may vary across sectors.

6 Conclusion

Ensuring the respectful and dignified treatment of humans when decisions that impact them are made by AI is critical for ensuring the ethical and human-centered deployment of the technology. This makes it important to understand how people view the appropriate use of AI technologies as they expand further into domains that directly impact people's lives and well-being, such as HRM contexts, and how AI decision making shapes people's perceptions of interactional justice. Our work takes a step in this direction and offers several future research paths to help ensure that the use of AI technologies now, and as they evolve in the future, will support feelings of respectful workplace treatment.

Declarations

• This study was funded by a grant received through the *Ethics in AI Research Initiative for the Asia Pacific*. This initiative is a partnership between Facebook and the Centre for Civil Society and Governance at The University of Hong Kong. The grant is in the form of an unrestricted gift, with the funding bodies having no influence over the conduct or reporting of any aspect of the research supported.
• The authors have no relevant financial or non-financial interests to disclose.
• The authors have no conflicts of interest to declare that are relevant to the content of this article.
• All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
• The authors have no financial or proprietary interests in any material discussed in this article.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods., 17*(4), 351–371.
Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society, 7*(2), 205395172094956.
Arasli, H., & Tümer, M. (2008). Nepotism, favoritism and cronyism: A study of their effects on job stress and job satisfaction in the banking industry of north Cyrus. *Social Behavior and Personality, 36*, 1237–1250.
Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*(3), 611–623.
Balasubramanian, S. (2021). The EU is proposing regulations on AI—and the impact on healthcare could be significant. *Forbes*. Retrieved from: https://www.forbes.com/sites/saibala/2021/04/25/the-eu-is-proposing-regulations-on-ai-and-the-impact-on-healthcare-could-be-significant/?sh=16cd73519be6
Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54*(4), 569–579.
Bastian, B., & Haslam, N. (2011). Experiencing dehumanization: Cognitive and emotional effects of everyday dehumanization. *Basic and Applied Social Psychology, 33*(4), 295–303.
Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*(3), 800–813.
Bies, R. (2001). *Interactional (in)justice In Greenberg & Cropanzano (Eds.), Advances in Organizational Justice* (pp. 89–118). Stanford University Press.

Bies, R. J., & Moag, J. S. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations, 1*, 43–55.

Binns, R. (2020). Human judgment in algorithmic loops: Individual justice and automated decision-making. *Regulation & Governance*. https://doi.org/10.1111/rego.12358

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & 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*, pp. 1–14. https://doi.org/10.1145/3173574.3173951.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41*(5), 545–547. https://doi.org/10.1188/14.ONF.545-547.

Christoff, K. (2014). Dehumanization in organizational settings: Some scientific and ethical considerations. *Frontiers in Psychology, 8,* 748. https://doi.org/10.3389/fnhum.2014.00748.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386.

Colquitt, J. A., & Rodell, J. B. (2015). Measuring justice and fairness. In: R. S. Cropanzano & M. L Ambrose (Eds.), *The oxford handbook of justice in the workplace* (p.187–202).Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199981410.013.8.

Colson, E. (2019). What AI-driven decision making looks like. *Harvard Business Review*. Retrieved from: https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like.

Cropanzano, R., Rupp, D. E., Mohler, C. J., & Schminke, M. (2001). Three roads to organizational justice. In J. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 20, pp. 1–113). Greenwich, CT: JAI.

Erdogan, B. (2002). Antecedents and consequences of justice perceptions in performance appraisals. *Human Resource Management Review, 12*(4), 555–578. https://doi.org/10.1016/S1053-4822(02)00070-0

European Commission. (2021). Regulatory framework proposal on Artificial Intelligence. Brussels, Belgium. Retrieved from: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288–299. https://doi.org/10.1037/0096-1523.1.3.288.

Guenole, N. & Feinzig, S. (2018). Competencies in the AI era. *IBM Smarter Workforce Institute*. Retrieved from: https://www.ibm.com/downloads/cas/ONNXK64Y

Gupta, M., Parra, C. M., & Dennehy, D. (2021). Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter? *Information Systems Frontiers,* 1–17. https://doi.org/10.1007/s10796-021-10156-2.

Hagras, H. (2018). Toward human-understandable, explainable AI. *Computer, 51*(9), 28–36.

Hamilton, V. L. (1978). Who is responsible? Toward a social psychology of responsibility attribution. *Social Psychology, 41*(4), 316–328. https://doi.org/10.2307/3033584.

Halloran, L., & Andrews, J. (2018). *Will you wait for the future to happen, or take a hand in shaping it? The future of work*. Ernst and Young.

Han, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann: Burlington.

IBM_CORP, R. (2021). *IBM SPSS Modeler for Windows, Version 1.82.1. IBM Corp*. NY: Armonk.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586. https://doi.org/10.1016/j.bushor.2018.03.007.

Karunakaran, A. (2018). *In cloud we trust? Normalization of uncertainties in online platform services*. Paper presented at the Academy of Management Proceedings.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174.

Körber, M. (2018). *Theoretical considerations and development of a questionnaire to measure trust in automation. In Proceedings 20th Triennial Congress of the IEA*. Springer. https://doi.org/10.31234/osf.io/nfc45.

Kumar, P., Dwivedi, Y., & Anand, A. (2021). Responsible artificial intelligence (AI) for value formation and market performance in healthcare. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10136-6

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*(2), 142–164. https://doi.org/10.1017/iop.2015.13.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1), 1–15. https://doi.org/10.1177/2053951718756684.

Lee M.K., Kusbit, D., Metsky, E., et al. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, South Korea*, pp. 1603–1612.

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction, 3(CSCW)*, Article 182. https://doi.org/10.1145/3359284.

Lind, E. (2001). Fairness heuristic theory: Justice judgments as pivotal cognitions in organizational relations. In J. Greenberg & R. Cropanzano (Eds.), *Advances in Organizational Justice* (pp. 56–88). Stanford University Press.

Lipshitz, R. (1989). Either a medal or a corporal: The effects of success and failure on the evaluation of decision making and decision makers. *Organizational Behavior and Human Decision Processes, 44*, 380–395.

Lockey, S., Gillespie, N., & Curtis, C. (2020). *Trust in artificial intelligence: Australian insights*. The University of Queensland and KPMG. https://doi.org/10.14264/b32f129.

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

Lucas, K. (2015). Workplace dignity: Communicating inherent, earned, and remediated dignity. *Journal of Management Studies, 52*(5), 621–646. https://doi.org/10.1111/joms.12133.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86.

Marr, B. (2018). The amazing ways Unilever uses artificial intelligence to recruit and train thousands of employees. *Forbes*. Retrieved from: https://www.forbes.com/sites/bernardmarr/2018/12/14/the-amazing-ways-how-unilever-uses-artificial-intelligence-to-recruit-train-thousands-of-employees/?sh=286750f56274

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*(1), 123–136. https://doi.org/10.1037/0021-9010.84.1.123.

Melick, S. R. (2020). *Development and validation of a measure of algorithm aversion*. Dissertation: Bowling Green State University.

McCarthy, J. (2007). From Here to Human-Level AI. *Artificial Intelligence, 171*(18), 1174–1182.

Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems, 2*(2), 1–25. https://doi.org/10.1145/1985347.1985353.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38.

Porra, J., Lacity, M., & Parks, M. (2020). Can Computer Based Human-Likeness Endanger Humanness? *Information Systems Frontiers, 22*(3), 533–547.

Pratt, M. G. (2009). Tips on writing up (and reviewing) qualitative research. *The Academy of Management Journal, 52*(5), 856–862. https://doi.org/10.5465/amj.2009.44632557.

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391.

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83. https://doi.org/10.1177/0008125619862257.

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology, 82*(3), 434–443. https://doi.org/10.1037/0021-9010.82.3.434.

Springer, A., & Whittaker, S. (2020). Progressive disclosure: When, why, and how do users want algorithmic transparency information? *ACM Transactions on Interactive Intelligent Systems, 10*(4), 1–32. https://doi.org/10.1145/3374218.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42. https://doi.org/10.1177/0008125619867910.

Ticona, J., & Mateescu, A. (2018). Trusted strangers: Carework platforms' cultural entrepreneurship in the on-demand economy. *New Media & Society, 20*(11), 4384–4404. https://doi.org/10.1177/1461444818773727.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38*(3), 505–520. https://doi.org/10.1016/j.ssresearch.2009.03.004.

Walsh, T., Levy, N., Bell, G., Elliott, A., Maclaurin, J., Mareels, I., & Wood, F. (2019). *The Effective and ethical development of Artificial Intelligence* (p. 250). ACOLA. https://acola.org/wp-content/uploads/2019/07/hs4_artificial-intelligence-report.pdf

Walton, D. (2004). A new dialectical theory of explanation. *Philosophical Explorations, 7*(1), 71–89. https://doi.org/10.1080/13869790320001868863.

Wang, W., Chen, L., Xiong, M., & Wang, Y. (2021). Accelerating AI adoption with responsible AI signals and employee engagement mechanisms in health care. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10154-4

Yassaee, M., & Mettler, T. (2019). Digital occupational health systems: What do employees think about it? *Information Systems Frontiers, 21*(4), 909–924.