When AI Enters the Workplace as Supervisors and Coworkers: Perceived Fairness in Human-AI Mixed Teams

I-CHEN CHIANG, National Chengchi University, Taiwan
YOYO TSUNG-YU HOU, National Chengchi University, Taiwan

Previous studies comparing algorithmic and human decision-making have produced inconsistent findings, largely focusing on individuals' responses to decision-makers. This study examines perceived fairness in workplace resource allocation by introducing comparison targets into a 2×2×2 between-subjects design manipulating supervisor type, coworker type, and allocation outcome. Results show that when comparison targets are made salient, individuals evaluate fairness based on whether decisions appear influenced by interpersonal favoritism, leading AI supervisors to be perceived as more impartial. Additionally, when collaborating with AI coworkers, participants tended to rationalize unequal allocations in terms of efficiency or performance, resulting in higher perceived fairness. These findings highlight the importance of "who" individuals compare themselves with, showing that perceived fairness in human-AI mixed teams is shaped by both decision-makers and comparison targets within a broader multi-agent context.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; **Collaborative and social computing**.

Additional Key Words and Phrases: AI decision-making, Perceived fairness, Human-AI collaboration

ACM Reference Format:
I-Chen Chiang and Yoyo Tsung-Yu Hou. 2026. When AI Enters the Workplace as Supervisors and Coworkers: Perceived Fairness in Human-AI Mixed Teams. In *Proceedings of AutomationXP26 Workshop of the 2026 CHI Conference on Human Factors in Computing Systems, April 14, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 8 pages.

1 Introduction

Algorithmic decision-making (ADM) is increasingly embedded in organizational contexts, such as recruitment, judicial sentencing, and donation allocation [6, 16],often with the promise of improving efficiency and reducing human bias. However, the growing adoption of ADM has also raised persistent concerns about fairness, particularly in decisions that directly affect individuals' outcomes.

Prior research comparing perceived fairness between human and AI decision-makers has yielded highly inconsistent findings. While some studies suggest that AI decisions are perceived as fairer than those made by humans, others report the opposite or no meaningful differences. These mixed results indicate that perceptions of algorithmic fairness are strongly context-dependent, making it difficult to draw generalized conclusions [25].

In workplace settings, resource allocation represents a common and highly consequential decision-making scenario. From an organizational perspective, understanding how individuals perceive fairness is critical, as perceived unfairness has been shown to trigger a range of negative outcomes, including increased turnover intentions, heightened hostility toward coworkers, and workplace aggression [8, 14]. As organizations increasingly consider delegating such decisions

Authors' Contact Information: I-Chen Chiang, National Chengchi University, Taipei, Taiwan, 113464034@nccu.edu.tw; Yoyo Tsung-Yu Hou, National Chengchi University, Taipei, Taiwan, yoyohou@nccu.edu.tw.

to AI systems, a key question emerges: whether algorithmic decision-making can enhance perceived fairness in these high-stakes contexts.

Moreover, prior research has primarily examined how decision-maker type and decision outcomes shape fairness perceptions, while largely overlooking a core principle of fairness theory: individuals often evaluate fairness through social comparison with others. In workplace contexts, coworkers serve as particularly salient referents for such comparisons. As AI systems are increasingly integrated into collaborative teams and function as coworkers, they may fundamentally reshape how individuals evaluate fairness.

To better understand algorithmic fairness in human-AI teams and promote effective collaboration, we conducted a 2x2x2 online experiment examining how supervisor type, coworker type, and allocation outcomes shape perceived fairness.

2 Related Work

2.1 Human versus Algorithmic Decision-Making Fairness

An emerging body of research has examined perceived fairness in algorithmic versus human decision-making. Some studies suggest that AI decision-makers are perceived as more fair than humans [3, 19]. These findings can be explained by the Machine Heuristic Model [26], which posits that individuals rely on mental shortcuts when interacting with machines, perceiving them as rule-governed, precise, accurate, objective, neutral, and infallible[27].

In contrast, other studies have found that human decision-makers are perceived as fairer than AI systems [4, 11, 30, 31]. Drawing on Mind Perception Theory [12, 13, 29], these studies argue that people generally attribute lower levels of agency and emotional experience to computational systems than to humans. As a result, AI decision-makers may be perceived as oversimplifying situational contexts and failing to account for complex interpersonal dynamics.

Additionally, some research indicates no significant differences in perceived fairness between human and algorithmic decisions [2]. These findings align with the Computers Are Social Actors (CASA) paradigm [21], which suggests that individuals tend to unconsciously apply social rules to computers or automated systems, leading to comparable psychological responses toward decisions made by humans and those made by AI.

Taken together, the inconsistency among previous findings appears to be closely related to differences in decision-making contexts, such as task characteristics[15] and task complexity[20], highlighting the complexity of fairness evaluations and the need for more diverse and nuanced research on ADM.

2.2 Fairness in Workplace Resource Allocation within HAI Teams

Organizations typically establish a bonus pool and allow managers or supervisors to distribute bonuses among employees to motivate higher levels of performance [18]. Individuals subject to these decisions tend to be highly sensitive to the distribution of resources, forming psychological reactions and fairness evaluations [5, 9, 17]. Prior research in organizational and HAI suggests that fairness judgments are largely outcome-driven and self-relevant. Individuals tend to perceive allocation decisions as fairer when the outcomes are favorable to themselves [7, 22, 24, 32]. Accordingly, we anticipate observing a similar effect in this study.

Hypothesis 1: Under an unfavorable allocation outcome (vs. favorable), perceived fairness will be lower. Beyond outcome favorability, Fairness Heuristic Theory suggests that when information is incomplete or ambiguous, individuals rely on salient cues, such as the characteristics of the decision-maker, to form fairness judgments [28].

In workplace resource allocation, where decision rules and processes are often opaque, the type of supervisor may therefore play a critical role in shaping perceived fairness.

RQ1 : How does supervisor type (AI vs. human) affect perceived fairness?

Equity theory, rooted in social comparison theory, suggests that individuals form judgments of fairness by comparing the ratio of their inputs to outcomes with those of others [1, 10]. In organizational settings, coworkers are among the most salient comparison targets. As AI coworkers are perceived differently from human coworkers in terms of agency and social presence [13], coworker type may play a critical role in shaping how allocation outcomes are interpreted and evaluated.

RQ2: How does coworker type (AI vs. human) affect perceived fairness?

3 Method

We conducted a 2x2x2 between-subjects experiment manipulating supervisor type(AI vs. human), coworker type (AI vs. human) and resource allocation outcome (favorable vs. unfavorable), comprising eight conditions.

3.1 Participants

We recruited adult participants through an online platform. After excluding those who failed the manipulation check or did not meet the age requirement, the final valid sample consisted of 128 participants. This study was conducted in accordance with institutional IRB rules. In terms of gender identity, 67.2% of the participants identified as female, 31.3% as male, and 1.5% as other. The mean age of the participants was 27.12 years (SD = 7.25). Regarding educational level, 67.2% held a bachelor's degree, 29.7% had a graduate degree or above, and 3.2% had a high school or vocational school education. Familiarity with AI was measured using a five-point Likert scale, with a mean score of 4.10 (SD = 0.71).

Table 1. Participant Distribution Across the Eight Conditions

		Coworker Type	
Allocation Outcome	Supervisor Type	AI	Human
Unfavorable	AI	16	21
	Human	16	15
Favorable	AI	20	14
	Human	16	10

3.2 Procedure

The experiment lasted approximately 10-12 minutes. Participants first read the study instructions and provided informed consent. They were then paired with a coworker (AI vs. human) and completed two collaborative tasks together. Following task completion, a bonus allocation outcome (favorable vs. unfavorable) was assigned by a supervisor (AI vs. human). Finally, participants completed a questionnaire assessing their subjective perceptions and were subsequently debriefed about the study.

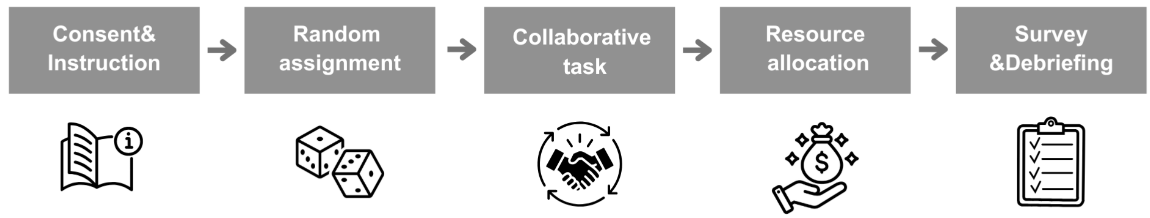

Fig. 1. Experimental procedure overview.

3.3 Materials

3.3.1 Manipulation of Independent Variables. Supervisor type (AI vs. human) and coworker type (AI vs. human) were manipulated through textual descriptions and visual icons. In addition, the resource allocation outcome was manipulated by informing participants that an additional bonus of NTD 100 would be distributed by the supervisor. In the favorable condition, participants received NTD 70 and their coworker received NTD 30; in the unfavorable condition, the allocation was reversed.

3.3.2 Collaboration task. To simulate a realistic workplace collaboration scenario, participants were asked to imagine working at a travel agency, where they partnered with a coworker to plan a customer's trip. This scenario was selected because it does not require specialized expertise and represents a familiar experience for most individuals. The study comprised two rounds of collaboration. In each round, one team member was responsible for booking the flight, while the other booked the hotel. Participants were informed that both roles were equally important and that responsibilities would be switched in the second round to ensure a balanced perception of contribution. During each task, a brief pause was introduced to simulate the coworker's thought process and decision-making. The coworker then asked for the participant's opinion, to which the participant was expected to respond. This design aimed to foster a sense of interaction and shared decision-making between the participant and the coworker.

3.4 Measures

Since workplace fairness research primarily focuses on distributive and procedural fairness, and given the manipulation constraints in HAI experiments, we adopted these two dimensions from the organizational justice framework to measure perceived fairness [7], which are applicable in HAI contexts [23]. All items were measured on a 5-point scale (1=to a small extent, 5=to a large extent). In addition, demographic information and open-ended questions were included to provide more comprehensive insights.

4 Preliminary results

This study employed a three-way Analysis of Variance(ANOVA) to examine differences in perceived fairness across the eight experimental conditions(2x2x2), with distributive fairness and procedural fairness analyzed separately.

4.1 Distributive Fairness

To test H1, allocation outcome had a strong effect on distributive fairness, with participants in the favorable condition perceiving significantly higher fairness than those in the unfavorable condition, $F(1, 120)=160.12$, $p<.001$, $\eta_p^2=0.573$. Addressing RQ1, supervisor type also influenced distributive fairness, such that AI supervisors were perceived as fairer

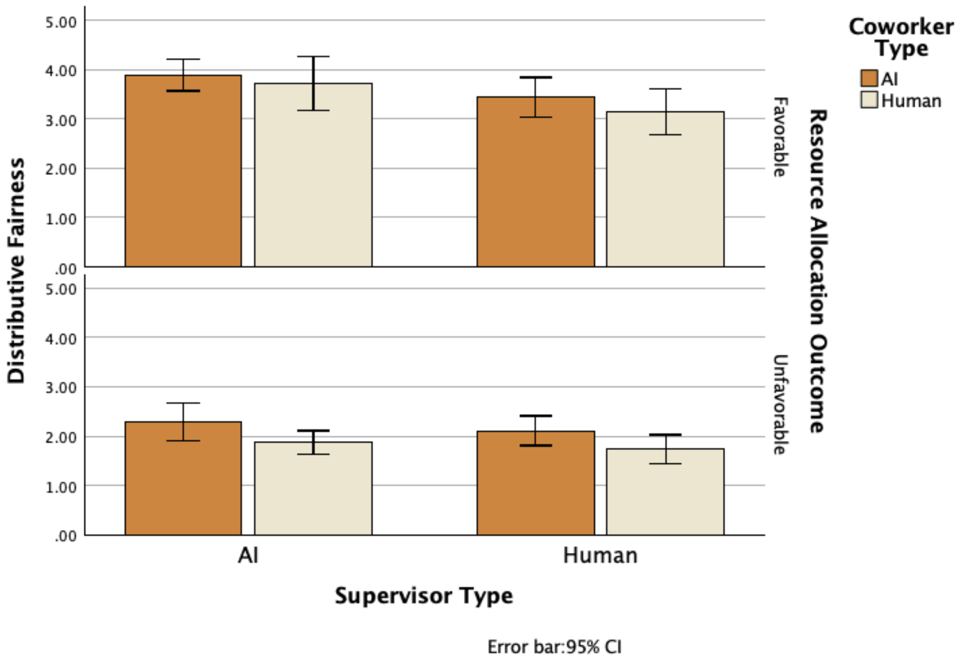

Fig. 2. Distributive Fairness in a Three-way ANOVA. Distributive fairness scores across supervisor type, coworker type, and allocation outcome. Favorable allocations were associated with higher perceived fairness than unfavorable allocations. AI supervisors and AI coworkers tended to be perceived as fairer than their human counterparts. Error bars represent 95% confidence intervals.

than human supervisors, $F(1, 120)=7.29$, $p= .008$, $\eta_p^2= 0.057$. Regarding RQ2, coworker type had a significant effect, with higher distributive fairness reported when collaborating with AI coworkers compared to human coworkers, $F(1,120)= 6.51$, $p =.012$, $\eta_p^2= 0.051$.

4.2 Procedural Fairness

Allocation outcome had a significant effect on procedural fairness, with participants in the favorable condition perceiving higher procedural fairness than those in the unfavorable condition, $F(1, 120)=32.74$, $p< .001$, $\eta_p^2= 0.21$. Supervisor type did not significantly influence procedural fairness($p=.127$). In contrast, coworker type had a significant effect, with procedural fairness perceived as higher when collaborating with AI coworkers than with human coworkers, $F(1,120)=5.286$, $p=.023$, $\eta_p^2= 0.042$.

5 Discussion

This study examined how supervisor type, coworker type, and allocation outcomes shape perceived fairness in HAI mixed teams. Consistent with prior research, fairness judgments were primarily driven by whether the allocation outcome was favorable to oneself, reinforcing the self-relevant nature of fairness perceptions in organizational contexts.

Beyond allocation outcomes, both supervisor type and coworker type further influenced perceived distributive fairness. When resources are limited and comparison targets are present, individuals may become especially sensitive to potential favoritism. Human supervisors are often perceived as susceptible to subjective judgment and interpersonal influence, whereas AI systems are viewed as more neutral and rule-based under the machine heuristic, leading to higher perceived fairness.

Regarding coworker type, participants collaborating with AI coworkers reported higher perceived fairness than those working with human coworkers. This effect may be explained by reduced interpersonal and moral concerns in social comparison. Compared with human coworkers, AI coworkers may not be perceived as equally legitimate or emotionally sensitive recipients of monetary rewards. As a result, when participants received more than an AI coworker, the outcome may have felt more justified and less morally problematic. At the same time, when participants received less, they were more likely to rationalize the unequal allocation in terms of the AI's higher efficiency or superior performance. As reflected in responses to open-ended questions, participants frequently attributed unequal outcomes to the AI's capabilities, which in turn increased perceived fairness.

These findings indicate that in workplace resource allocation contexts, considering who individuals are allocated resources alongside plays a critical role in shaping fairness perceptions, as well as in how the fairness of decision-makers is evaluated. AI agents, whether acting as decision-makers or as collaborative coworkers, demonstrate the potential to enhance perceived fairness. This study advances research on algorithmic fairness by emphasizing the importance of multi-agent comparison processes and provides practical insights for organizations considering the adoption of AI in human-AI collaboration.

6 Limitations and Future Work

Several limitations should be acknowledged, pointing to directions for future research. As an exploratory study, the findings should be interpreted with caution. Future work could employ larger samples to increase statistical power and improve the generalizability of the results. Additionally, the current design focused on short-term and relatively minimal interaction between team members. Future research may examine longer-term collaboration, richer forms of coworker interaction, or repeated allocation scenarios to better understand how fairness perceptions develop and stabilize over time in HAI teams. Future studies could explore additional contextual factors, such as varying levels of decision transparency or task complexity, to further clarify the boundary conditions under which AI agents enhance or undermine perceived fairness in organizational settings.

7 Conclusion

Through an online experiment, this study shows that comparison targets in multi-agent decision contexts shape fairness evaluations. AI supervisors may reduce perceived favoritism and enhance fairness perceptions, while AI coworkers may prompt individuals to rationalize allocation outcomes, thereby increasing perceived fairness.

By extending prior binary comparisons between humans and AI to a multi-agent perspective, this work offers a more comprehensive and ecologically valid account of workplace decision-making, with implications for how AI is positioned within collaborative organizational settings.

References

[1] J Stacy Adams. 1963. Towards an understanding of inequity. *The journal of abnormal and social psychology* 67, 5 (1963), 422.

[2] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & society* 35, 3 (2020), 611–623.

[3] Bing Bai, Hengchen Dai, Dennis J Zhang, Fuqiang Zhang, and Haoyuan Hu. 2022. The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments. *Manufacturing & Service Operations Management* 24, 6 (2022), 3060–3078.

[4] Pinar Barlas, Styliani Kleanthous, Kyriakos Kyriakou, and Jahna Otterbacher. 2019. What makes an image tagger fair?. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*. 95–103.

[5] Kathryn M Bartol and Abhishek Srivastava. 2002. Encouraging knowledge sharing: The role of organizational reward systems. *Journal of leadership & organizational studies* 9, 1 (2002), 64–76.

[6] Fei Cai, Jiashu Zhang, and Lei Zhang. 2024. The impact of artificial intelligence replacing humans in making human resource management decisions on fairness: A case of resume screening. *Sustainability* 16, 9 (2024), 3840.

[7] Jason A Colquitt, Donald E Conlon, Michael J Wesson, Christopher OLH Porter, and K Yee Ng. 2001. Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of applied psychology* 86, 3 (2001), 425.

[8] Joerg Dietz, Sandra L Robinson, Robert Folger, Robert A Baron, and Martin Schulz. 2003. The impact of community violence and an organization's procedural justice climate on workplace aggression. *Academy of Management Journal* 46, 3 (2003), 317–326.

[9] James H Dulebohn and Joseph J Martocchio. 1998. Employee perceptions of the fairness of work group incentive pay plans. *Journal of Management* 24, 4 (1998), 469–488.

[10] Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.

[11] Nicholas Folger, Prisca Brosi, Jutta Stumpf-Wollersheim, and Isabell M Welpe. 2022. Applicant reactions to digital selection methods: A signaling perspective on innovativeness and procedural justice. *Journal of Business and Psychology* 37, 4 (2022), 735–757.

[12] Heather M Gray, Kurt Gray, and Daniel M Wegner. 2007. Dimensions of mind perception. *science* 315, 5812 (2007), 619–619.

[13] Kurt Gray and Daniel M Wegner. 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition* 125, 1 (2012), 125–130.

[14] Jerald Greenberg and R Cropanzano. 1993. The social side of fairness: Interpersonal and informational classes of organizational justice. *Justice in the workplace: Approaching fairness in human resource management* (1993).

[15] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big data & society* 5, 1 (2018), 2053951718756684.

[16] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. 2017. A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3365–3376.

[17] Gerald S Leventhal. 1976. The distribution of rewards and resources in groups and organizations. In *Advances in experimental social psychology*. Vol. 9. Elsevier, 91–131.

[18] Victor S Maas, Marcel Van Rinsum, and Kristy L Towry. 2012. In search of informed discretion: An experimental investigation of fairness and trust reciprocity. *The Accounting Review* 87, 2 (2012), 617–644.

[19] Frank Marcinkowski, Kimon Kieslich, Christopher Starke, and Marco Lünich. 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*. 122–130.

[20] Rosanna Nagtegaal. 2021. The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government information quarterly* 38, 1 (2021), 101536.

[21] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[22] David T Newman, Nathanael J Fast, and Derek J Harmon. 2020. When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes* 160 (2020), 149–167.

[23] Jessica Ochmann, Leonard Michels, Verena Tiefenbeck, Christian Maier, and Sven Laumer. 2024. Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. *Information Systems Journal* 34, 2 (2024), 384–414.

[24] Robert E Ployhart and Ann Marie Ryan. 1997. Toward an explanation of applicant reactions: An examination of organizational justice and attribution frameworks. *Organizational behavior and human decision processes* 72, 3 (1997), 308–335.

[25] Christopher Starke, Janine Baleis, Birte Keller, and Frank Marcinkowski. 2022. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society* 9, 2 (2022), 20539517221115189.

[26] S Shyam Sundar et al. 2008. *The MAIN model: A heuristic approach to understanding technology effects on credibility*. MacArthur Foundation Digital Media and Learning Initiative Cambridge, MA.

[27] S Shyam Sundar and Jinyoung Kim. 2019. Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI Conference on human factors in computing systems*. 1–9.

[28] Kees Van den Bos, E Allan Lind, and Henk AM Wilke. 2012. The psychology of procedural and distributive justice viewed from the perspective of fairness heuristic theory. In *Justice in the workplace*. Psychology Press, 49–66.

[29] Adam Waytz and Michael I Norton. 2014. Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. *Emotion* 14, 2 (2014), 434.

[30] Magdalena Wojcieszak, Arti Thakur, João Fernando Ferreira Gonçalves, Andreu Casas, Ericka Menchen-Trevino, and & Miriam Boon. [n. d.]. Can AI enhance people's support for online moderation and their openness to dissimilar political views? *Journal of Computer-Mediated Communication* ([n. d.]).

[31] Jang Wonseok, Kang Young Woo, and Kang Yeonheung. 2021. Who made the decisions: Human or robot umpires? The effects of anthropomorphism on perceptions toward robot umpires. *Telematics and Informatics* 64 (2021), 101695.

[32] Gizem Yalcin, Sarah Lim, Stefano Puntoni, and Stijn MJ van Osselaer. 2022. Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research* 59, 4 (2022), 696–717.