Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 19 August 2024
Revised 13 September 2024
Accepted 26 September 2024
Published online 01 October 2024

Intan. Sari[1], Agus. Santoso[2*]

1 Department of Child and Family Studies, Universitas Padjadjaran, Bandung, Indonesia
2 Department of Educational Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia

Corresponding author email address:
agus.santoso@ugm.ac.id

How to cite this article:
Sari, I., & Santoso, A . (2024). Perceived AI Fairness and Organizational Commitment: The Mediating Role of Psychological Safety. *Future of Work and Digital Management Journal, 2*(4), 44-54. https://doi.org/10.61838/fwdmj.2.4.5

Perceived AI Fairness and Organizational Commitment: The Mediating Role of Psychological Safety

ABSTRACT

This study aimed to investigate the relationship between perceived AI fairness and organizational commitment, with psychological safety examined as a potential mediating variable. A descriptive correlational design was employed using a sample of 443 employees from public and private sector organizations in Indonesia. The sample size was determined using the Morgan and Krejcie table for large populations. Standardized instruments were used to measure organizational commitment (Organizational Commitment Questionnaire by Mowday et al., 1979), psychological safety (Psychological Safety Scale by Edmondson, 1999), and perceived AI fairness (Perceived Fairness in Algorithmic Decision-Making Scale by Lee, 2018). Data were analyzed using SPSS-27 for descriptive and Pearson correlation statistics, and AMOS-21 for structural equation modeling (SEM). Model fit was assessed using established indices including χ^2/df, CFI, TLI, GFI, AGFI, and RMSEA. Perceived AI fairness was significantly and positively correlated with organizational commitment ($r = .39$, $p < .001$) and psychological safety ($r = .51$, $p < .001$). Psychological safety was also positively correlated with organizational commitment ($r = .47$, $p < .001$). SEM results confirmed good model fit ($\chi^2/df = 2.47$, CFI = 0.96, RMSEA = 0.057) and demonstrated that perceived AI fairness significantly predicted psychological safety ($\beta = 0.51$, $p < .001$), which in turn predicted organizational commitment ($\beta = 0.39$, $p < .001$). The indirect effect of AI fairness on commitment through psychological safety was also significant ($\beta = 0.20$, $p < .001$), indicating partial mediation. The findings suggest that perceived fairness in AI systems contributes to higher organizational commitment, both directly and through enhanced psychological safety. Organizations should consider both technological transparency and supportive interpersonal environments to foster employee trust and engagement in AI-integrated workplaces.

Keywords: Perceived AI Fairness, Psychological Safety, Organizational Commitment.

Introduction

As artificial intelligence (AI) systems increasingly permeate organizational decision-making processes, concerns about their fairness, transparency, and ethical use have grown in tandem. Particularly in human resources functions—such as recruitment, performance evaluation, and promotion—algorithmic tools are deployed to optimize efficiency and consistency. However, these applications raise critical questions about how AI systems are perceived by employees and whether perceptions of fairness influence key organizational outcomes, including commitment and retention. As organizations strive to maintain a motivated and loyal workforce amidst growing technological mediation, it becomes vital to investigate how perceptions of AI fairness impact employees' psychological experiences and behaviors within the workplace [1-3].

AI fairness refers to the degree to which algorithmic decisions are perceived as equitable, transparent, and free from bias [4]. Despite the promise of objectivity and impartiality, AI systems often reflect and perpetuate existing societal inequalities

due to biased data sets, flawed design assumptions, or opaque processing mechanisms [5, 6]. Research has shown that when employees view algorithmic processes as unfair or inscrutable, their trust in organizational leadership diminishes, resulting in disengagement and reduced performance. Conversely, when AI systems are perceived as fair—demonstrating consistency, explainability, and accountability—employees are more likely to accept automated decisions and sustain organizational loyalty [7, 8]. Thus, perceived AI fairness emerges as a critical antecedent to positive employee outcomes in the digital workplace.

Organizational commitment, defined as the psychological attachment and loyalty that employees feel toward their employer, is a well-documented predictor of workplace outcomes including reduced turnover, higher job performance, and increased citizenship behaviors [9]. However, commitment is not formed in a vacuum; rather, it is shaped by the contextual and interpersonal dynamics within the organization. As AI systems assume a more central role in managing human capital, employee perceptions of fairness and psychological safety become key predictors of how deeply they identify with their organization. In this sense, the interplay between AI fairness and psychological mechanisms—such as trust and safety—may either reinforce or erode organizational commitment, particularly in environments characterized by digital transformation and constant technological adaptation [10, 11].

Psychological safety, a term originally coined by Edmondson, refers to a shared belief that the workplace is safe for interpersonal risk-taking. In psychologically safe environments, employees feel comfortable expressing concerns, asking questions, and admitting mistakes without fear of humiliation or retaliation. Psychological safety has consistently been linked to team learning, innovation, and employee engagement [12, 13]. In the context of AI-enabled workplaces, psychological safety may serve as a critical buffer between the technological environment and employees' attitudinal responses. If employees perceive AI systems as opaque, unaccountable, or discriminatory, it may erode their sense of control and safety, which in turn undermines their commitment to the organization [14, 15].

Recent studies have emphasized the mediating role of psychological safety in shaping employee reactions to digital technologies and automated decision-making systems. For instance, when employees feel safe to question the logic or fairness of AI-generated outcomes, they are more likely to engage with these tools constructively, leading to higher organizational trust and identification [16, 17]. Conversely, environments lacking in psychological safety may cause employees to suppress dissent, disengage from decision-making processes, and develop a sense of alienation. As such, the experience of psychological safety may explain the psychological pathway through which perceived AI fairness translates into organizational commitment [18, 19].

Theoretically, this study builds on existing frameworks of fairness theory and organizational behavior by integrating concepts from psychological safety literature. Fairness theory posits that individuals assess the legitimacy of decisions based on procedural justice (the fairness of the process), distributive justice (the fairness of outcomes), and interactional justice (the respectfulness of communication) [2, 5]. Psychological safety, in turn, functions as a moderator of how employees interpret and respond to potentially threatening situations, including algorithmic assessments and decisions. When the organization fosters a culture of openness and support, it may neutralize the alienating effects of AI technologies and enhance the perceived legitimacy of AI use [20, 21]. This aligns with findings that psychological safety strengthens positive organizational climates, mediates role stress, and promotes ethical cultures—especially in contexts characterized by high uncertainty [22, 23].

In healthcare, education, and judicial sectors—where algorithmic tools are increasingly deployed—scholars have underscored the dual-edged nature of AI adoption. While automation can enhance objectivity and efficiency, it can also depersonalize interactions and obscure decision-making accountability, especially when communication lacks human empathy or feedback mechanisms [3, 24]. These concerns are magnified in organizations with low psychological safety, where employees may hesitate to challenge algorithmic outcomes or suggest improvements. On the contrary, a psychologically safe climate empowers employees to engage in critical reflection, articulate doubts, and co-develop adaptive strategies that reinforce organizational commitment [10, 12].

From a managerial perspective, the integration of AI fairness and psychological safety offers a practical framework for enhancing employee retention and trust. Organizational leaders can no longer assume that technological efficiency alone will secure commitment; rather, they must invest in transparent communication, participatory governance, and an inclusive climate that encourages critical engagement with digital tools [4, 9]. This includes explaining the rationale behind algorithmic decisions, offering recourse mechanisms, and ensuring that all employees feel safe voicing their perspectives—even when they challenge prevailing norms or outputs [8, 13]. Empirical studies suggest that such practices not only mitigate the risks of digital alienation but also foster a stronger alignment between individual and organizational values [7, 14].

Despite growing interest in the psychological dimensions of AI adoption, limited empirical work has directly examined the mediating role of psychological safety in the relationship between perceived AI fairness and organizational commitment—particularly in diverse, non-Western contexts. Most existing literature is concentrated in North America and Europe, with insufficient attention paid to emerging economies where digital transformation is rapidly evolving but often constrained by regulatory gaps, cultural differences, and infrastructure disparities [17, 18]. For example, in Indonesia—where this study is situated—the accelerated adoption of AI in both public and private sectors raises urgent questions about ethical implementation, employee perceptions, and institutional trust. This cultural and organizational complexity necessitates a more nuanced exploration of how AI systems are internalized and contested by employees in diverse workplace settings [1, 6].

To address these gaps, the present study investigates the extent to which perceived AI fairness predicts organizational commitment, and whether this relationship is mediated by psychological safety.

Methods and Materials

Study Design and Participants

This study employed a descriptive correlational design to investigate the relationship between perceived AI fairness and organizational commitment, with psychological safety as a mediating variable. The target population consisted of employees from various private and public organizations in Indonesia. A total of 443 participants were selected using stratified random sampling, based on the Morgan and Krejcie (1970) sample size determination table for large populations. Participants were recruited across sectors including technology, education, healthcare, and finance to ensure diversity in organizational contexts where algorithmic decision-making is present.

Data Collection

To assess organizational commitment, the study employed the Organizational Commitment Questionnaire (OCQ) developed by Mowday, Steers, and Porter (1979). This widely used instrument consists of 15 items rated on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree), designed to measure the affective attachment of employees to their organization. The OCQ focuses primarily on affective commitment, reflecting the extent to which individuals identify with and are involved in the organization. The overall score is obtained by averaging the item responses, with higher scores indicating stronger organizational commitment. Numerous studies have confirmed the construct validity and internal consistency of the OCQ, with Cronbach's alpha values typically exceeding 0.85, supporting its reliability across diverse organizational contexts.

Psychological safety was measured using the Psychological Safety Scale developed by Amy Edmondson (1999). This tool contains 7 items, each rated on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree), and assesses the extent to which individuals perceive their team environment as safe for interpersonal risk-taking. The scale is unidimensional and does not include subscales; rather, it provides a single composite score by averaging the items, where higher scores indicate greater perceived psychological safety. The scale has been widely used in organizational behavior research, and its reliability has been consistently supported, with Cronbach's alpha values typically above 0.70. Its validity has also been confirmed in multiple empirical studies linking psychological safety with team learning, voice behavior, and performance outcomes.

To evaluate perceived AI fairness, this study utilized the Perceived Fairness in Algorithmic Decision-Making Scale developed by Lee (2018). This measure comprises 12 items grouped into three subscales: procedural fairness (e.g., transparency and consistency of the AI process), distributive fairness (e.g., perceived equity in outcomes), and interactional fairness (e.g., respectful and informative communication about AI decisions). Each item is rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree), and subscale scores can be computed alongside an overall fairness score. The tool has demonstrated good construct validity and internal reliability, with reported Cronbach's alpha values between 0.75 and 0.88 across subscales. It has been increasingly adopted in recent research exploring trust and acceptance of AI systems in organizational contexts.

Data analysis

Data analysis was conducted using SPSS version 27 and AMOS version 21. Descriptive statistics (frequencies, percentages, means, and standard deviations) were computed for demographic variables. To examine the relationships between the dependent variable (organizational commitment) and the independent variables (perceived AI fairness and psychological safety), Pearson correlation coefficients were calculated. Furthermore, to assess the mediating role of psychological safety, a Structural Equation Modeling (SEM) approach was used. Model fit indices such as the Chi-square/df ratio, RMSEA, CFI, and TLI were evaluated to determine the adequacy of the proposed model.

Findings and Results

Of the 443 participants, 257 (58.0%) were female and 186 (42.0%) were male. The participants' ages ranged from 21 to 58 years, with 129 individuals (29.1%) aged 21–30, 176 (39.7%) aged 31–40, 91 (20.5%) aged 41–50, and 47 (10.6%) aged above 50. Regarding educational background, 112 participants (25.3%) held a diploma, 201 (45.4%) held a bachelor's degree, and

130 (29.3%) held a master's degree or higher. In terms of employment sector, 104 (23.5%) were employed in technology, 98 (22.1%) in education, 121 (27.3%) in healthcare, and 120 (27.1%) in finance and related sectors. The majority of participants (276 individuals, 62.3%) reported having experience with AI-based systems in their workplace.

Table 1

Descriptive Statistics for Study Variables (N = 443)

Variable	Mean	Standard Deviation
Organizational Commitment	84.72	10.46
Psychological Safety	27.39	4.81
Perceived AI Fairness	42.16	6.29

The descriptive statistics presented in Table 1 show that the mean score for organizational commitment was 84.72 (SD = 10.46), suggesting a moderately high level of commitment among employees. The mean score for psychological safety was 27.39 (SD = 4.81), indicating that participants generally perceived their work environment as moderately safe for interpersonal risk-taking. The mean score for perceived AI fairness was 42.16 (SD = 6.29), which reflects a moderate to high perception of fairness in algorithmic decision-making processes within their organizations.

Prior to conducting the Pearson correlation and SEM analysis, the assumptions of normality, linearity, homoscedasticity, and multicollinearity were examined. Skewness and kurtosis values for all continuous variables were within the acceptable range of ±2, indicating approximate normality (e.g., skewness = 0.51 for organizational commitment; kurtosis = 1.17 for perceived AI fairness). Scatterplots confirmed the linear relationship between each independent variable and the dependent variable. Homoscedasticity was supported by visual inspection of residual plots. Multicollinearity was ruled out, with variance inflation factor (VIF) values ranging between 1.03 and 1.28, well below the threshold of 10. These results supported the suitability of the data for correlation and structural equation modeling.

Table 2

Pearson Correlations Among Study Variables (N = 443)

Variable	1	2	3
1. Organizational Commitment	—		
2. Psychological Safety	.47** (p < .001)	—	
3. Perceived AI Fairness	.39** (p < .001)	.51** (p < .001)	—

As shown in Table 2, organizational commitment was significantly and positively correlated with psychological safety ($r = .47$, $p < .001$) and perceived AI fairness ($r = .39$, $p < .001$). Additionally, perceived AI fairness was positively associated with psychological safety ($r = .51$, $p < .001$). These findings support the theoretical premise that AI fairness and psychological safety are both influential predictors of employee commitment.

Table 3

Fit Indices for the Structural Equation Model

Fit Index	Value	Acceptable Threshold
Chi-Square (χ^2)	241.73	—
Degrees of Freedom (df)	98	—
χ^2/df	2.47	< 3.00
GFI	0.93	≥ 0.90
AGFI	0.91	≥ 0.90
CFI	0.96	≥ 0.95
TLI	0.95	≥ 0.95
RMSEA	0.057	≤ 0.08

Table 3 presents the model fit indices for the structural equation model. The Chi-square value was 241.73 with 98 degrees of freedom, and the χ^2/df ratio was 2.47, which is below the threshold of 3.00, indicating a good fit. Other indices also met or exceeded recommended cutoffs: GFI = 0.93, AGFI = 0.91, CFI = 0.96, TLI = 0.95, and RMSEA = 0.057. These results suggest that the hypothesized model provides a good fit to the observed data.

Table 4

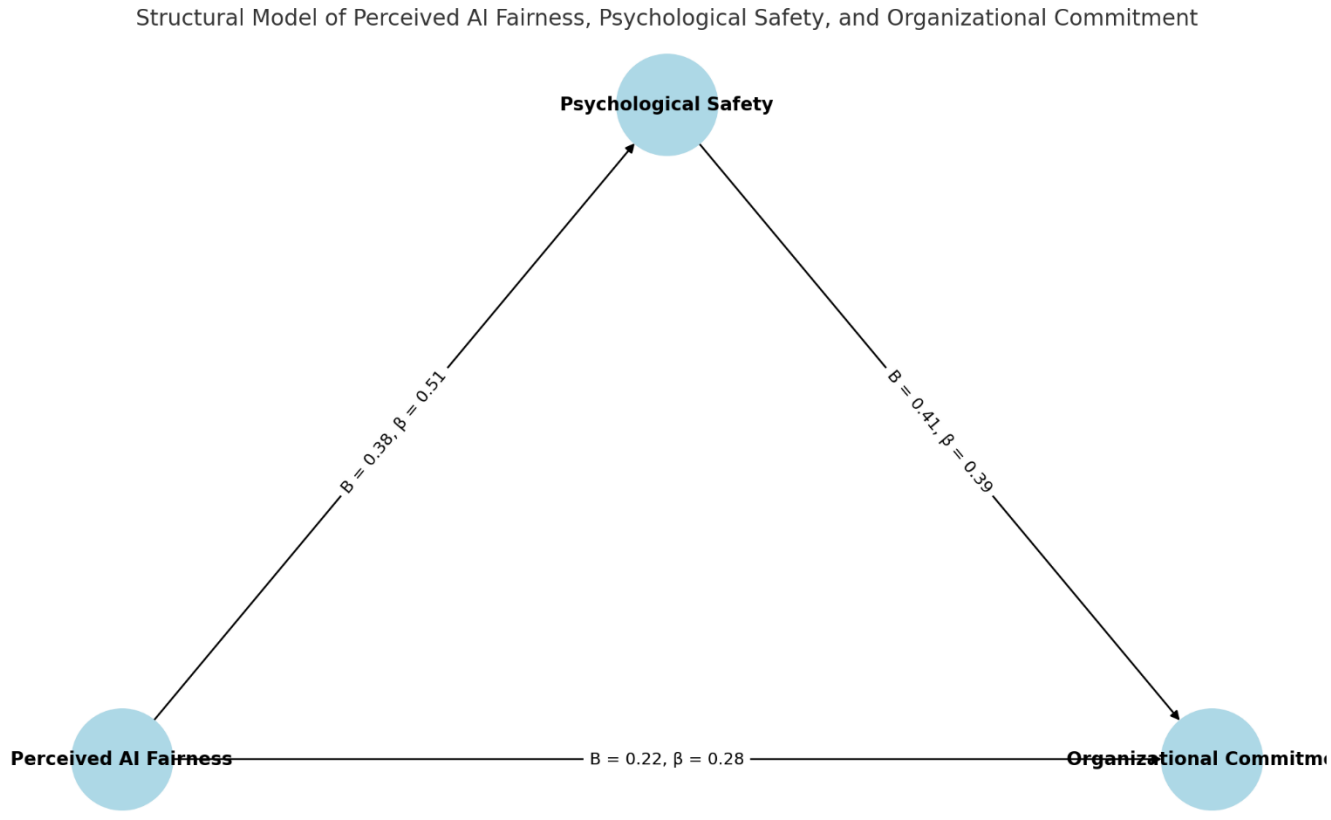
Total, Direct, and Indirect Effects in the Structural Model

Path	B	S.E.	Beta	p
AI Fairness → Psychological Safety	0.38	0.05	0.51	<.001
AI Fairness → Organizational Commitment	0.22	0.06	0.28	<.001
Psychological Safety → Org. Commitment	0.41	0.07	0.39	<.001
AI Fairness → Org. Commitment (Indirect)	0.16	0.04	0.20	<.001
AI Fairness → Org. Commitment (Total)	0.38	0.05	0.48	<.001

Table 4 illustrates the path coefficients for the direct, indirect, and total effects within the structural model. Perceived AI fairness had a significant positive effect on psychological safety (B = 0.38, β = 0.51, p < .001), and both perceived AI fairness (B = 0.22, β = 0.28, p < .001) and psychological safety (B = 0.41, β = 0.39, p < .001) significantly predicted organizational commitment. The indirect effect of perceived AI fairness on organizational commitment through psychological safety was also significant (B = 0.16, β = 0.20, p < .001), indicating a partial mediation. The total effect of AI fairness on organizational commitment (direct + indirect) was 0.38 (β = 0.48, p < .001), confirming the model's overall strength and coherence.

Figure 1

Model with Beta Values



Discussion and Conclusion

The findings of this study reveal significant associations between perceived AI fairness, psychological safety, and organizational commitment among employees in Indonesia. Correlation analysis demonstrated that both perceived AI fairness and psychological safety were positively associated with organizational commitment. Furthermore, structural equation modeling (SEM) confirmed that psychological safety significantly mediates the relationship between perceived AI fairness and organizational commitment, supporting the hypothesized model. These results suggest that when employees view AI systems as fair and equitable, they are more likely to feel psychologically secure within their organizational environment, which in turn enhances their emotional attachment and loyalty to the organization.

The positive direct relationship observed between perceived AI fairness and organizational commitment aligns with the theoretical expectation that fairness perceptions in technological systems significantly influence how employees relate to their organizations. This finding is supported by prior research that highlights the importance of algorithmic transparency, procedural justice, and equitable outcomes in shaping user trust and satisfaction in AI applications [1, 2]. When employees perceive that AI systems are used to make decisions impartially and with accountability, it fosters a sense of respect and recognition, which enhances organizational identification and commitment. For instance, Ho et al. (2023) emphasized that perceptions of fairness in AI-driven judicial and administrative decisions contribute to public confidence and procedural legitimacy, concepts that are transferable to organizational dynamics [3].

Moreover, the mediating role of psychological safety underscores the psychological mechanisms through which perceptions of fairness influence employee attitudes. Psychological safety is increasingly recognized as a foundational element for trust and open communication in organizations, especially in technologically complex and fast-evolving environments [12, 22]. In this study, psychological safety was found to significantly mediate the fairness-commitment relationship, implying that fairness alone may not directly generate commitment unless it is accompanied by a climate that allows employees to express themselves without fear. This is consistent with the findings of Kang (2024), who demonstrated that perceived psychological safety amplifies the positive impact of procedural systems on risk perception and workplace behavior [14].

The study's results also confirm that environments perceived as psychologically safe strengthen the internalization of fair practices. When employees feel safe to question decisions, challenge norms, or report concerns, they are more likely to accept and even support the presence of AI in the workplace. This aligns with Porter-Stransky et al. (2024), who found that departmental interventions promoting psychological safety improved collaboration and openness, even in highly structured and hierarchical environments [16]. Similarly, Hunt et al. (2021) showed that in mental health services, increased psychological safety correlated with higher staff engagement and reduced resistance to procedural innovations [15]. These findings reinforce the notion that psychological safety is not merely a passive state but an active catalyst that enables organizational practices—such as AI adoption—to take root and be sustained.

Several studies in both organizational and educational contexts lend additional support to the mediating function of psychological safety. Atiku et al. (2023) emphasize the necessity of fostering safety within family businesses to promote team

effectiveness and continuity [9]. Likewise, Ganuzin and Kovaleva (2024) found that when pedagogical leaders emphasize emotional safety and inclusivity, educational environments become more innovative and responsive [17]. Applying these findings to corporate settings suggests that psychological safety can act as a protective layer that mitigates the uncertainty and ambiguity often associated with algorithmic processes.

Interestingly, the present study also aligns with literature that critiques assumptions of uniform psychological safety within teams or organizations. Loignon and Wormington (2022) warned that assuming shared psychological safety can mask power dynamics and suppress dissent, particularly among junior employees or marginalized groups [20]. This suggests that even when AI fairness is perceived as high, without intentional efforts to promote inclusivity and dialogue, organizations may fall short of achieving meaningful employee commitment. The current findings thus echo the view that fairness and safety are deeply interrelated constructs that must be cultivated simultaneously and strategically [11, 24].

Furthermore, the positive relationship between AI fairness and psychological safety corroborates studies examining ethical AI deployment and human-centered design. Nguyen et al. (2023) explored how content moderators' perceptions of explanation fairness in hate speech detection influenced their psychological comfort and perceived control [5]. In organizational settings, a similar logic applies: when employees understand how AI systems arrive at decisions and trust that those decisions are not biased or arbitrary, their sense of safety is preserved or even enhanced. Ethical AI frameworks, such as those proposed by Verma et al. (2023), emphasize fairness as a precursor to psychological security, particularly in high-stakes environments such as autonomous systems and digital surveillance [2].

Additionally, the study contributes to a broader discourse on AI integration in emerging economies, such as Indonesia. As AI tools are increasingly embedded into HR, operations, and management practices across Southeast Asia, concerns about transparency, worker alienation, and ethical oversight remain salient [6, 18]. The present findings suggest that organizations operating in such transitional contexts must attend not only to the technical performance of AI but also to how these tools are socially and psychologically experienced by employees. Psychological safety becomes especially critical in non-Western work cultures, where hierarchical structures and cultural norms may inhibit open discussion of technology-related concerns [19, 21].

The implications of this study also align with findings in healthcare and legal domains where AI use is expanding rapidly. Ünver and Asan (2022) explored how patient safety in AI-driven healthcare systems is contingent on both trust in the technology and the communication strategies of health professionals [4]. Similarly, in judicial contexts, Sung (2024) documented how perceptions of AI fairness influence both compliance and morale in criminal justice systems [7]. These sectoral parallels suggest that regardless of context, the psychological implications of AI implementation require careful management to avoid unintended consequences on workforce well-being and morale.

Finally, the results contribute to the growing recognition that psychological safety is an essential condition for responsible innovation. As noted by Zeglin et al. (2024), psychological safety not only reduces distress but also creates a space for critical engagement and ethical reflection in times of rapid change [10]. Therefore, psychological safety is not only a mediator in this model but also a strategic asset that organizations must actively nurture to navigate the complexities of digital transformation. As organizations strive to integrate AI in a way that aligns with their values and missions, fostering fair, transparent, and safe environments becomes not just desirable but necessary.

Despite its contributions, this study is not without limitations. First, the cross-sectional design limits causal inference; while associations between variables were identified, the directionality of influence cannot be definitively established. Longitudinal studies are needed to assess how perceptions of AI fairness and psychological safety evolve over time and impact organizational commitment. Second, the reliance on self-report measures may introduce common method bias, as participants' responses might have been influenced by social desirability or subjective interpretation. Third, although the sample was drawn from various sectors, it was geographically limited to Indonesia, which may affect the generalizability of the findings to other cultural or organizational settings. Finally, while SEM offers robust insights into the relationships among variables, other mediating or moderating factors—such as organizational culture, leadership style, or individual resilience—were not examined in this study and may play important roles.

Future research should consider adopting longitudinal or experimental designs to strengthen causal conclusions and assess changes over time in psychological safety and organizational commitment. Researchers are also encouraged to examine potential moderators, such as demographic characteristics, digital literacy, or organizational transparency, which might influence the strength of the observed relationships. Cross-cultural studies comparing employees in different national or regional contexts would also enrich our understanding of how cultural values shape responses to AI fairness and psychological safety. Moreover, qualitative approaches such as interviews or focus groups could uncover deeper insights into employees' lived experiences with algorithmic systems and their nuanced perceptions of fairness and trust. Expanding the model to include additional outcomes such as job satisfaction, voice behavior, or turnover intentions could also enhance its explanatory power.

Organizations adopting AI technologies should prioritize transparency and fairness in algorithmic decision-making by providing clear explanations and opportunities for employee feedback. Creating psychologically safe environments where employees feel empowered to question and discuss AI-driven outcomes can significantly enhance trust and commitment. Training programs that foster digital literacy, ethical awareness, and inclusive communication can further support employee adaptation and engagement. Leaders must actively model openness and responsiveness to concerns about AI systems and involve employees in discussions about their design and implementation. Investing in both technological infrastructure and human-centered organizational practices will ensure that AI deployment contributes not only to operational efficiency but also to a positive and sustainable work culture.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] J. Li and M. Chignell, "FMEA-AI: AI Fairness Impact Assessment Using Failure Mode and Effects Analysis," *Ai and Ethics*, vol. 2, no. 4, pp. 837-850, 2022, doi: 10.1007/s43681-022-00145-9.

[2] S. Verma, P. Pali, M. Dhanwani, and S. Jagwani, "Ethical AI: Developing Frameworks for Responsible Deployment in Autonomous Systems," *Ijmrset*, vol. 6, no. 04, pp. 1-14, 2023, doi: 10.15680/ijmrset.2023.0604036.

[3] Y. J. Ho, W. Jabr, and Y. Zhang, "Ai Enforcement: Examining the Impact of Ai on Judicial Fairness and Public Safety," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4533047.

[4] M. B. Ünver and O. Asan, "Role of Trust in AI-Driven Healthcare Systems: Discussion From the Perspective of Patient Safety," *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 11, no. 1, pp. 129-134, 2022, doi: 10.1177/2327857922111026.

[5] T. Nguyen, J. Xu, A. Roy, H. Daumé, and M. Carpuat, "Towards Conceptualization of "Fair Explanation": Disparate Impacts of Anti-Asian Hate Speech Explanations on Content Moderators," 2023, doi: 10.18653/v1/2023.emnlp-main.602.

[6] H. Yadav, H. Sinha, A. Manhar, and A. Patle, "Ethical Integration of Artificial Intelligence in Criminology: Addressing Challenges for a Safer Society," *International Research Journal of Modernization in Engineering Technology and Science*, 2023, doi: 10.56726/irjmets41779.

[7] Y.-E. Sung, "Criminal Psychology and Artificial Intelligence (AI): Risk Factors and Implications," *Korea Assoc Crim Psychol*, vol. 20, no. 4, pp. 105-130, 2024, doi: 10.25277/kcpr.2024.20.4.105.

[8] Y. Chen, "Pandoras Pixel Box: The Rise of AI Art and the Ethical Dilemma of Creativity," *Lecture Notes in Education Psychology and Public Media*, vol. 28, no. 1, pp. 135-143, 2023, doi: 10.54254/2753-7048/28/20231315.

[9] S. O. Atiku, F. A. Soneye, and R. Anane-Simon, "Fostering Psychological Safety for Team Effectiveness in Family Businesses," pp. 84-121, 2023, doi: 10.4018/978-1-6684-8748-8.ch004.

[10] A. Zeglin, M. L. McCarthy, S. Nedorost, M. Dell, and L. Logio, "Psychological Distress in the Era of Psychological Safety," *Journal of General Internal Medicine*, vol. 40, no. 1, pp. 38-40, 2024, doi: 10.1007/s11606-024-08893-6.

[11] Y. Zhang and M. Wan, "The Double-Edged Sword Effect of Psychological Safety Climate: A Theoretical Framework," *Team Performance Management*, vol. 27, no. 5/6, pp. 377-390, 2021, doi: 10.1108/tpm-01-2021-0005.

[12] I. Goller and J. Bessant, "Practising Psychological Safety," pp. 133-146, 2023, doi: 10.4324/9781003226178-10.

[13] N. Jamal, V. N. Young, J. Shapiro, M. Brenner, and C. E. Schmalbach, "Patient Safety/Quality Improvement Primer, Part IV: Psychological Safety—Drivers to Outcomes and Well-being," *Otolaryngology*, vol. 168, no. 4, pp. 881-888, 2022, doi: 10.1177/01945998221126966.

[14] L. Kang, "Describing the Impact of Psychological Safety on Risk Prevention: A Threshold Model Construction," *Work*, vol. 79, no. 1, pp. 277-288, 2024, doi: 10.3233/wor-230234.

[15] D. F. Hunt, J. Bailey, B. Lennox, M. Crofts, and C. Vincent, "Enhancing Psychological Safety in Mental Health Services," *International Journal of Mental Health Systems,* vol. 15, no. 1, 2021, doi: 10.1186/s13033-021-00439-1.

[16] K. A. Porter-Stransky, K. J. Horneffer, L. Bauler, K. Gibson, C. M. Haymaker, and M. Rothney, "Improving Departmental Psychological Safety Through a Medical School-Wide Initiative," *BMC Medical Education,* vol. 24, no. 1, 2024, doi: 10.1186/s12909-024-05794-4.

[17] V. M. Ganuzin and E. Kovaleva, "Development of Pedagogical Competencies in Creating a Psychologically Safe Educational Environment," pp. 164-168, 2024, doi: 10.46727/c.v1.21-22-03-2024.p164-168.

[18] S. Sanduleac, "Psychological Security Versus Psychological Safety," *Moldoscopie,* no. 2(97), pp. 100-106, 2023, doi: 10.52388/1812-2566.2022.2(97).09.

[19] T. L. Khudyakova, Y. V. Klepach, and M. R. Arpentieva, "Components and Criteria for the Psychological Safety of Education," *Society and Security Insights,* vol. 5, no. 4, pp. 155-170, 2023, doi: 10.14258/ssi(2022)4-09.

[20] A. C. Loignon and S. V. Wormington, "Psychologically Safe for Some, but Not All? The Downsides of Assuming Shared Psychological Safety Among Senior Leadership Teams," 2022, doi: 10.35613/ccl.2022.2048.

[21] J. E. Swain, K. Conkey, Y. Kalkstein, and O. Strauchler, "Exploring the Utility of Psychological Safety in the Armed Forces," *Journal of Character and Leadership Development,* vol. 11, no. 2, pp. 55-67, 2024, doi: 10.58315/jcld.v11.288.

[22] B. Fyhn, H. Bang, T. E. Sverdrup, and V. Schei, "Safe Among the Unsafe: Psychological Safety Climate Strength Matters for Team Performance," *Small Group Research,* vol. 54, no. 4, pp. 439-473, 2022, doi: 10.1177/10464964221121273.

[23] K. S. Vogt, J. Baker, M. Morys-Edge, S. Kendal, E. Mizen, and J. Johnson, "'I Think the First Priority Is Physically Safe First, Before You Can Actually Get Psychologically Safe': Staff Perspectives on Psychological Safety in Inpatient Mental Health Settings," *Journal of Psychiatric and Mental Health Nursing,* vol. 32, no. 2, pp. 276-287, 2024, doi: 10.1111/jpm.13101.

[24] Y. V. Smyk and A. Y. Kachimskaya, "Teacher's Capacity for Psychological Safety as a Condition for Schoolchildren's Psychological Safety," *Science for Education Today,* vol. 11, no. 1, pp. 42-58, 2021, doi: 10.15293/2658-6762.2101.03.