

# Human-Centered Artificial Intelligence in Human Resource Management: The Effect of Perceived AI Fairness and Algorithmic Trust on Work Engagement

Ahmad Shalahuddin<sup>1\*</sup>, Nur Afifah<sup>2</sup>

<sup>1,2</sup>Faculty of Economic and Business, Universitas Tanjungpura, Pontianak

---

## ABSTRACT

This study examines whether Perceived AI Fairness is associated with Work Engagement through Algorithmic Trust in a Human-Centered Artificial Intelligence in Human Resource Management context. A quantitative cross-sectional survey involved 100 employees in Pontianak who reported direct exposure to an AI-enabled HR process containing automated scoring, prediction, ranking, natural-language interaction, or algorithmic recommendation; employees exposed only to ordinary administrative digitization were not eligible. Data were analyzed using PLS-SEM. Perceived AI Fairness was positively associated with Algorithmic Trust and Work Engagement, while Algorithmic Trust was positively associated with Work Engagement. The positive direct association and the statistically significant indirect pathway are consistent with complementary mediation, but they do not establish a causal transmission mechanism because the data are cross-sectional and self-reported. Recommendations concerning explainability, algorithmic audits, appeal mechanisms, meaningful human review, preservation of dignity, and limits on algorithmic authority are presented as theory-informed governance implications rather than as constructs directly tested by the survey.

---

---

## ARTICLE INFO

### Keywords:

Perceived AI Fairness;  
Algorithmic Trust;  
Work Engagement;  
Human-Centered AI;  
Human Resource  
Management.

### Article Information:

Received: 18/06/2026

Revised: 05/10/2026

Accepted: 07/10/2026

### ISSN:

2985-3168 (Online)

2985-3222 (Print)

---

\*Corresponding Author at: Ahmad Shalahuddin

Department of Management, Faculty of Economics and Business, Universitas Tanjungpura, Pontianak

E-mail address: [ahmad.shalahuddin@ekonomi.untan.ac.id](mailto:ahmad.shalahuddin@ekonomi.untan.ac.id)<sup>1</sup>

The work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)



## **1. Introduction**

Work engagement is a positive psychological state that reflects employees' energy, dedication, and cognitive-affective involvement in their work, making it an important indicator of the effectiveness of contemporary Human Resource Management practices (Borst et al., 2020; Decuypere & Schaufeli, 2020). In digital organizational contexts, engagement is influenced not only by job design and leadership, but also by the way intelligent technologies, particularly Artificial Intelligence, shape employees' work experiences, perceived control, and sense of meaningfulness at work (Kwon & Kim, 2020; Parker & Grote, 2022). AI in HRM can function as a job resource when it supports decision-making, reduces administrative burdens, and enhances personalized employee experiences (Budhwar et al., 2022; Vrontis et al., 2022). However, AI can also become a new job demand when algorithmic decisions are perceived as opaque, difficult to understand, or threatening to human autonomy in the workplace (Kellogg et al., 2020; Jarrahi et al., 2021). Therefore, work engagement in AI-HRM should be understood through the lens of human-centered AI, which emphasizes fairness, trust, safety, and human dignity (Shneiderman, 2020; Raisch & Krakowski, 2021).

The main phenomenon underlying this study is the growing use of AI in recruitment, selection, performance appraisal, talent analytics, training, and HR decision-making processes that were previously dominated by human judgment (Budhwar et al., 2022; Vrontis et al., 2022). Although AI-HRM offers efficiency, consistency, and stronger analytical capability, its implementation also raises risks related to algorithmic bias, decision opacity, excessive surveillance, and weakened perceptions of fairness among employees (Köchling & Wehner, 2020; Langer & Landers, 2021). Previous studies show that individuals tend to evaluate AI-based systems more critically when decisions lack explanation, provide no opportunity for appeal, or reduce meaningful human involvement (Acikgoz et al., 2020; Newman et al., 2020). Thus, the urgency of this study emerges from the need to examine whether perceived AI fairness is associated with work engagement through a statistically compatible indirect pathway involving algorithmic trust.

Perceived AI fairness refers to the extent to which employees believe that AI systems in HRM operate fairly, accurately, consistently, transparently, non-discriminatorily, and respectfully toward individuals affected by algorithmic decisions (Köchling & Wehner, 2020; Starke et al., 2022). This perception is important because AI-based HRM decisions may influence selection, promotion, compensation, performance appraisal, and employees' career development (Acikgoz et al., 2020; Newman et al., 2020). Algorithmic fairness should not be assessed only through technical accuracy, but also through data representativeness, procedural clarity, and the availability of corrective mechanisms when AI-based decisions are perceived as harmful (Langer & Landers, 2021; Shin, 2021). From a human-centered AI perspective, AI fairness signals that the organization continues to protect employees' voice, rights, and dignity in technology-based decision-making processes (Shneiderman, 2020; Raisch & Krakowski, 2021).

Algorithmic trust is defined as employees' belief that AI systems or organizational algorithms possess competence, reliability, integrity, transparency, and acceptable intentions in supporting workplace decisions (Glikson & Woolley, 2020; Jacovi et al., 2021). In AI-HRM, trust is crucial because employees often cannot directly observe the technical logic of AI systems, so they assess trustworthiness through cues such as fairness, consistency, explainability, and organizational accountability (Araujo et al., 2020; Shin, 2021). Perceived AI fairness is theoretically expected to be positively associated with algorithmic trust because

fairer evaluations may correspond with lower uncertainty and stronger beliefs that the system does not operate arbitrarily (Glikson & Woolley, 2020; Choung et al., 2023). In turn, stronger algorithmic trust may correspond with employees' ability to direct energy, attention, and dedication toward their work; the cross-sectional design, however, does not establish the temporal direction of these relationships (Kwon & Kim, 2020; Parker & Grote, 2022).

The novelty of this research lies in its focus on the "contested terrain" of algorithmic control (Kellogg et al., 2020). Unlike conventional organizational justice, algorithmic fairness demands technical explainability and formal appeal mechanisms. We theorize that fair AI-HRM acts as a 'job resource' (based on the Job Demands-Resources model) that fosters vigor and dedication, provided it is perceived as an augmentative tool rather than an invasive surveillance mechanism. By bridging organizational justice and AI governance, this study uniquely positions algorithmic trust not as a generic outcome, but as a distinct psychological state rooted in a system's ethical and technical competence.

Although AI-HRM can improve efficiency and analytical capability, technological benefits do not automatically produce employee acceptance (Budhwar et al., 2022; Vrontis et al., 2022). Existing research has emphasized HR transformation, algorithmic management, and automated decision-making, while the mechanism linking perceived fairness, trust in AI, and work engagement remains insufficiently integrated (Kellogg et al., 2020; Starke et al., 2022; Meijerink & Bondarouk, 2023). This study addresses that gap by linking organizational justice and the Job Demands-Resources model to explain employees' psychological responses to AI-enabled HR decisions.

Based on the above discussion, this study aims to examine the association between perceived AI fairness and work engagement, including whether Algorithmic Trust forms a statistically compatible indirect pathway between them in the context of Human-Centered Artificial Intelligence in Human Resource Management (Shneiderman, 2020; Budhwar et al., 2022). Theoretically, this study contributes by integrating human-centered AI, organizational justice, algorithmic trust, and work engagement perspectives to explain employees' psychological responses to AI-HRM systems (Köchling & Wehner, 2020; Glikson & Woolley, 2020; Kwon & Kim, 2020). Empirically, the study tests only Perceived AI Fairness, Algorithmic Trust, and Work Engagement. Explainability, auditability, appeal mechanisms, meaningful human review, preservation of dignity, and limits on algorithmic authority are therefore treated as theory-informed governance implications derived from the broader Human-Centered AI literature rather than as separate empirical findings of the survey.

### **Theoretical Framework and Hypothesis Development**

Organizational justice is the principal framework for explaining how employees evaluate AI-enabled HR decisions. Procedural justice concerns the consistency, neutrality, accuracy, voice, and correctability of decision procedures, whereas distributive justice concerns the perceived fairness of outcomes. Technical explainability is conceptually separate: it is a system property that can supply information for a justice judgment but is not itself equivalent to fairness. The Job Demands-Resources model complements this framework by explaining the motivational consequence of that judgment. A fair and contestable AI-HRM process can operate as a job resource by reducing uncertainty and preserving employee voice; an opaque or coercive process can become a job demand that consumes psychological resources (Kellogg et al., 2020; Parker & Grote, 2022; Starke et al., 2022).

Fair procedures and outcomes signal that the organization and its AI system act consistently, avoid arbitrary treatment, and permit correction. These justice cues reduce uncertainty and strengthen employees' trust beliefs and willingness to rely on algorithm-supported HR decisions (Glikson & Woolley, 2020; Bankins et al., 2022). H1: Perceived AI Fairness is positively associated with Algorithmic Trust.

When employees perceive AI-enabled HR decisions as fair, they are more likely to view the technology as a supportive organizational resource rather than a threat to autonomy or career security. This positive appraisal should preserve energy, dedication, and involvement in work (Parker & Grote, 2022). H2: Perceived AI Fairness is positively associated with Work Engagement.

Algorithmic Trust represents confidence in the system's competence, reliability, integrity, and dependable operation together with a willingness to rely on its recommendations under appropriate human oversight. Trust reduces vigilance and uncertainty and allows employees to direct psychological resources toward their work roles. H3: Algorithmic Trust is positively associated with Work Engagement.

Organizational justice is also expected to show an indirect association with engagement through Algorithmic Trust because fair AI-enabled procedures may correspond with stronger trust beliefs, which in turn may be associated with greater engagement. The proposed pathway does not exclude a remaining direct association between fairness and engagement, and its cross-sectional estimation should not be interpreted as evidence of temporal causal mediation. H4: Algorithmic Trust is statistically consistent with an indirect positive pathway linking Perceived AI Fairness and Work Engagement.

## **2. Methodology**

This study uses a quantitative explanatory-predictive approach with a cross-sectional survey design to examine associations among Perceived AI Fairness, Algorithmic Trust, and Work Engagement. Because all variables were measured once from the same employees, the design does not establish temporal precedence or causality. Perceived AI Fairness is the predictor, Algorithmic Trust is the mediator, and Work Engagement is the outcome in the proposed model (Bankins et al., 2022; Starke et al., 2022).

The target population comprised employees in Pontianak organizations that used AI-enabled HRM. For this study, an AI-enabled HR system was defined as a system that performs at least one of the following functions: machine-learning prediction, automated scoring or ranking, algorithmic recommendation, natural-language interaction, or automated decision support affecting an HR outcome. Ordinary digital recordkeeping, payroll, document storage, or attendance capture without these functions was classified as digital HR rather than AI-HRM and did not independently satisfy eligibility. A priori power planning used a conservative medium effect size ( $f^2 = 0.15$ ), two predictors for the most complex endogenous construct,  $\alpha = .05$ , and power = .80. The minimum required sample was 68; the achieved sample of 100 exceeded this criterion and yields approximately .94 power under the same assumptions. PLS-SEM was selected for its predictive orientation and mediation model, not merely because the sample was small (Hair & Alamer, 2022).

Purposive sampling was applied. Eligible respondents had to work in Pontianak and report direct exposure to at least one clearly described AI-enabled HR function involving automated scoring, prediction, ranking, recommendation, natural-language processing, or algorithmic decision support. Respondents whose experience was limited to ordinary digital

attendance or administrative HR software were excluded. Exposure had to concern a personally experienced or consequential HR process, such as recruitment, performance appraisal, training recommendation, promotion, prediction, or another employee decision. The retained manuscript dataset available for this revision does not contain recoverable identifiers for the number of organizations represented, respondents per organization, industry composition, or a frequency distribution of AI-HRM application types. These characteristics therefore cannot be reconstructed without the original respondent-level dataset. This limitation is important because perceptions may differ across AI-HRM applications and clustering of multiple respondents within the same organization could reduce observation independence. PLS-SEM with SmartPLS was used to assess the measurement and structural models. The retained output supports reporting of loadings, Cronbach's alpha, composite reliability, AVE, HTMT point estimates, path coefficients,  $R^2$ , and bootstrap significance. The original SmartPLS project is required to complete rho\_A, bootstrapped HTMT confidence intervals, outer and inner VIF,  $f^2$  by structural path,  $Q^2/PLSpredict$ , SRMR, and bias-corrected bootstrap confidence intervals for direct, indirect, and total effects. No values for these diagnostics are reconstructed or inferred in this revision because doing so would create unsupported results.

Perceived AI Fairness was operationalized as an overarching evaluative judgment regarding the fairness of an AI-enabled HR decision. Its seven indicators capture procedural consistency, non-discrimination, outcome fairness, relevance of decision data, procedural clarity, opportunity to question the result, and availability of correction. The reflective specification is theoretically defensible to the extent that these items are treated as manifestations of a broader perceived-fairness judgment rather than as independent design features that mechanically form fairness. This specification is also consistent with the study's focus on employees' global evaluation of the legitimacy of AI-HRM rather than on estimating separate justice dimensions. Nevertheless, the indicator content spans procedural, distributive, informational, and voice/correctability elements that are conceptually distinguishable. Accordingly, the present unidimensional reflective model should be interpreted cautiously, especially for H1, H2, and H4, and future validation should compare it with a multidimensional or higher-order specification using the original item-level data. Technical explainability is treated as a contextual system characteristic that can inform fairness judgments, not as identical to fairness itself. Algorithmic Trust was adapted to the AI-HRM context from Choung et al. (2023) and measured as trust beliefs and willingness to rely on an AI-supported HR process; competence, reliability, integrity, security, and reliance were assessed, while transparency was treated as an antecedent cue rather than trust itself. Work Engagement used nine items representing vigor, dedication, and absorption. All items used a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Appendix A reports the item wording and construct assignment.

Common method risk was addressed procedurally by defining the focal AI-HRM experience before the construct questions, separating predictor, mediator, and outcome blocks, using neutral instructions, assuring respondents that no individual answer would be shared with employers, and avoiding language that suggested a preferred response. Statistical evidence beyond the retained measurement output is still required: a full-collinearity VIF assessment or theoretically unrelated marker-variable test should be reported from the original dataset before final submission. The cross-sectional model may also be affected by reverse causality because highly engaged employees could judge AI systems more favorably.

For this reason, coefficients are interpreted as predictive associations, not causal effects; longitudinal, experimental, or instrumental-variable designs are needed to evaluate temporal ordering and endogeneity.

### 3. Result and Discussion

#### Result

The analysis included 100 employees in Pontianak who passed the AI-exposure criterion: each reported direct experience with an HR system performing automated scoring, prediction, ranking, natural-language interaction, algorithmic recommendation, or decision support. Ordinary digital attendance or administrative software alone did not qualify. The available retained data do not provide recoverable counts for the number of organizations represented, respondents per organization, industry composition, firm size, digital maturity, or the distribution of AI-HRM application types. Consequently, the pooled analysis cannot determine whether responses are independent across organizations or whether the structural relationships differ between recruitment, appraisal, training, promotion, prediction, or recommendation applications. Respondents were predominantly aged 26–35 years (46%), followed by 36–45 years (28%), 18–25 years (19%), and above 45 years (7%). Men represented 52% and women 48%. Most held a bachelor's degree (67%), followed by diploma (18%), master's degree (10%), and senior high school or equivalent (5%). Tenure was 1–5 years for 43%, 6–10 years for 31%, less than one year for 14%, and more than ten years for 12%.

The descriptive statistical analysis shows that *Perceived AI Fairness* obtained a mean score of 4.06 with a standard deviation of 0.52, indicating that respondents generally perceived the use of AI in HRM practices as fair. This perception was mainly reflected in the consistency of procedures, relevance of data, and transparency of AI-based decisions. *Algorithmic Trust* recorded a mean score of 3.96 with a standard deviation of 0.55, suggesting that employees' trust in algorithmic systems was relatively high, although improvements are still needed in terms of system explainability and accountability. Meanwhile, *Work Engagement* obtained the highest mean score of 4.11 with a standard deviation of 0.49, indicating that employees demonstrated a strong level of engagement in terms of vigor, dedication, and absorption in their work.

**Table 1.** Descriptive Statistics of Research Variables

| Variable              | Mean | Standard Deviation | Category |
|-----------------------|------|--------------------|----------|
| Perceived AI Fairness | 4.06 | 0.52               | High     |
| Algorithmic Trust     | 3.96 | 0.55               | High     |
| Work Engagement       | 4.11 | 0.49               | High     |

The measurement model was evaluated using outer loading, Cronbach's Alpha, Composite Reliability, and Average Variance Extracted. The results show that all indicators had outer loading values above 0.70, indicating that all indicators were valid. The Perceived AI Fairness construct obtained a Cronbach's Alpha value of 0.913, Composite Reliability of 0.931, and AVE of 0.660. The Algorithmic Trust construct obtained a Cronbach's Alpha value of 0.891, Composite Reliability of 0.920, and AVE of 0.697. Furthermore, the Work Engagement construct obtained a Cronbach's Alpha value of 0.902, Composite Reliability of 0.925, and AVE of 0.674. Therefore, all constructs fulfilled the criteria for convergent validity and internal consistency reliability.

Discriminant validity assessment using the heterotrait–monotrait ratio showed that all reported point estimates were below 0.85. HTMT was 0.719 between Perceived AI Fairness and Algorithmic Trust, 0.606 between Perceived AI Fairness and Work Engagement, and 0.692 between Algorithmic Trust and Work Engagement. These point estimates support empirical distinctiveness among the three constructs. Bootstrapped HTMT confidence intervals were not contained in the retained output and must be added from a rerun before the requested complete assessment can be claimed.

The exact indicator counts were seven for Perceived AI Fairness, five for Algorithmic Trust, and nine for Work Engagement. Cronbach's alpha, composite reliability, AVE, and HTMT point estimates support the retained measurement model at the level reported in the available output. However, the Round-3 diagnostics requested by the reviewer—rho\_A, bootstrapped HTMT confidence intervals, outer and inner VIF,  $f^2$  by structural path,  $Q^2/PLSpredict$ , SRMR, and bias-corrected bootstrap confidence intervals for direct, indirect, and total effects—are not contained in the retained output. Because these statistics cannot be validly reconstructed from the summary coefficients, the manuscript does not claim that the extended PLS-SEM diagnostic assessment is complete. A final submission that fully satisfies this diagnostic request requires rerunning the original SmartPLS project and replacing the diagnostic status table with the resulting estimates and confidence intervals.

**Table 2.** Status of Additional PLS-SEM Diagnostics Requested in Round 3

| Diagnostic           | Reporting status                                                                             |
|----------------------|----------------------------------------------------------------------------------------------|
| rho_A                | Not available in retained output; rerun original SmartPLS project.                           |
| HTMT inference       | 95% bootstrap CI not available in retained output; rerun required.                           |
| Collinearity         | Outer and inner VIF not available in retained output; rerun required.                        |
| Effect size          | $f^2$ by structural path not available in retained output; rerun required.                   |
| Predictive relevance | $Q^2/PLSpredict$ not available in retained output; rerun required.                           |
| Model fit            | SRMR not available in retained output; rerun required.                                       |
| Bootstrap inference  | Bias-corrected 95% CI for direct, indirect, and total effects not available; rerun required. |

**Table 3.** HTMT Discriminant Validity Test

| Construct Relationship                    | HTMT Value | Interpretation |
|-------------------------------------------|------------|----------------|
| Perceived AI Fairness – Algorithmic Trust | 0.719      | Valid          |
| Perceived AI Fairness – Work Engagement   | 0.606      | Valid          |
| Algorithmic Trust – Work Engagement       | 0.692      | Valid          |

The structural model supports the four hypotheses at the reported significance level, but the coefficients are interpreted as cross-sectional associations rather than causal effects. Perceived AI Fairness was positively associated with Algorithmic Trust ( $\beta = 0.725$ ,  $t = 13.248$ ,  $p < .001$ ), supporting H1. Algorithmic Trust was positively associated with Work Engagement ( $\beta = 0.525$ ,  $t = 5.831$ ,  $p < .001$ ), supporting H3. Perceived AI Fairness was also positively associated with Work Engagement ( $\beta = 0.282$ ,  $t = 2.973$ ,  $p = .003$ ), supporting H2. Algorithmic Trust had  $R^2 = 0.526$ , while Work Engagement had  $R^2 = 0.579$ . These estimates indicate predictive relationships within the specified model. Bias-corrected bootstrap confidence intervals are required from the original SmartPLS output before complete inferential reporting can be claimed.

**Table 4.** Hypothesis Testing Results

| Variable Relationship                       | Coefficient | T-Statistic | P-Value | Result       |
|---------------------------------------------|-------------|-------------|---------|--------------|
| Perceived AI Fairness<br>→Algorithmic Trust | 0.725       | 13.248      | < .001  | H1 supported |
| Algorithmic Trust → Work<br>Engagement      | 0.525       | 5.831       | < .001  | H3 supported |
| Perceived AI Fairness → Work<br>Engagement  | 0.282       | 2.973       | 0.003   | H2 supported |

The indirect association of Perceived AI Fairness with Work Engagement through Algorithmic Trust was positive and significant ( $\beta = 0.381$ ,  $t = 5.274$ ,  $p < .001$ ), supporting H4. The direct association remained positive and significant ( $\beta = 0.282$ ,  $p = .003$ ), and the reported total effect was 0.663. The common positive direction of the direct and indirect paths is statistically consistent with complementary mediation. This terminology describes the pattern of associations in the fitted model; it does not demonstrate that Algorithmic Trust causally transmits an effect of Perceived AI Fairness onto Work Engagement. Reverse, reciprocal, and omitted-variable explanations remain possible because all constructs were measured once from the same respondents. Bias-corrected confidence intervals for the indirect, direct, and total effects remain necessary from the original SmartPLS bootstrap output for complete mediation inference.

**Table 5.** Mediation Test of Algorithmic Trust

| Mediation Relationship                                            | Indirect Effect | T-Statistic | P-Value | Direct Effect        | Interpretation          |
|-------------------------------------------------------------------|-----------------|-------------|---------|----------------------|-------------------------|
| Perceived AI Fairness →<br>Algorithmic Trust →<br>Work Engagement | 0.381           | 5.274       | < .001  | 0.282 ( $p = .003$ ) | Complementary mediation |

## Discussion

Perceived AI Fairness was positively and significantly associated with Algorithmic Trust. Employees who reported higher fairness perceptions also tended to report higher trust in the AI systems used in HRM. In this study, AI fairness reflects employees' judgments concerning procedural consistency, relevant data use, non-discrimination, procedural clarity, voice, correctability, and fair outcomes. The coefficient of 0.725 indicates a strong positive association within the specified model, but it should not be interpreted as evidence that fairness causes trust. This pattern is consistent with Glikson and Woolley (2020), who argued that human trust in artificial intelligence is related to perceived system competence, reliability, transparency, and integrity. In HRM contexts, such trust is especially consequential because algorithm-supported decisions may concern performance appraisal, promotion, training opportunities, compensation, and career development. Because Perceived AI Fairness was modeled as one reflective construct despite conceptually heterogeneous domains, interpretation of H1 should remain at the level of an overall fairness perception rather than implying that each justice dimension contributes identically.

The association is also consistent with Köchling and Wehner (2020), who emphasized that algorithmic systems in HRM require careful attention to fairness because algorithms may reproduce or amplify bias when data are unrepresentative or decision procedures are inadequately monitored. The present data show that employees' fairness judgments and algorithmic trust covary positively; they do not establish that trust emerges as a causal

consequence of fairness. Employees may evaluate AI not only from a technical perspective but also according to whether the system appears legitimate, accountable, and respectful. Alternative temporal ordering is also plausible, because employees who already trust organizational technology may evaluate the same AI-HRM process more favorably. These possibilities reinforce the need for longitudinal or experimental designs.

Algorithmic Trust was positively and significantly associated with Work Engagement ( $\beta = 0.525$ ). Employees reporting higher trust in AI-HRM also tended to report greater vigor, dedication, and absorption. This relationship is compatible with the Job Demands–Resources perspective, in which a trusted technology may be appraised as a supportive resource rather than a source of uncertainty or threat (Parker & Grote, 2022). However, the cross-sectional results do not show that increasing trust will necessarily increase engagement, because engagement itself or other organizational conditions may shape trust evaluations. The coefficient therefore represents a predictive association within the fitted model rather than a causal effect.

From the perspective of the Job Demands–Resources model, fair and trustworthy AI-HRM may function as a job resource when employees experience the system as supportive, intelligible, and compatible with meaningful human oversight. AI systems can potentially reduce administrative burdens, provide feedback, or support personalized development, whereas opaque or coercive systems may be experienced as job demands associated with surveillance, uncertainty, and reduced autonomy (Kellogg et al., 2020; Parker & Grote, 2022). These theoretical mechanisms help interpret the observed associations, but they were not independently manipulated or measured as causal processes in this cross-sectional survey. Accordingly, the results are consistent with, rather than proof of, the proposition that human-centered AI conditions support engagement.

Perceived AI Fairness was also positively and significantly associated with Work Engagement ( $\beta = 0.282$ ) after Algorithmic Trust was included in the model. This remaining direct association indicates that employees who judged AI-HRM as fair also tended to report stronger engagement independently of the modeled indirect pathway. The result is compatible with organizational justice reasoning that fair treatment, employee voice, and respect for dignity are related to positive work attitudes (Bankins et al., 2022). It does not demonstrate that fairness directly produces engagement, and interpretation of H2 should also acknowledge that the seven fairness indicators span conceptually distinct justice-related domains. A higher-order specification may clarify whether particular fairness dimensions show different relationships with engagement.

Algorithmic Trust shows a complementary mediating pattern in the statistical association between Perceived AI Fairness and Work Engagement. Both the indirect pathway through trust and the remaining direct pathway are positive and significant. This pattern is statistically compatible with an indirect mechanism in which fairness, trust, and engagement are linked in the theorized sequence, but the cross-sectional data cannot establish temporal precedence or causal transmission. Reverse and reciprocal relationships remain plausible, and the interpretation of H4 is further conditioned by the current unidimensional reflective specification of Perceived AI Fairness. Accordingly, the mediation result should be read as model-consistent indirect association rather than evidence that Algorithmic Trust causally carries an effect from fairness to engagement.

Theoretically, this study contributes to the development of the human-centered artificial intelligence perspective in HRM. Shneiderman (2020) emphasized that human-centered AI should be reliable, safe, transparent, and respectful of human dignity. In this study, these principles are reflected in the importance of perceived fairness and algorithmic trust. Human-centered AI-HRM does not merely pursue speed and accuracy in decision-making. It also ensures that employees understand the decision process, have opportunities to question AI-based outcomes, and continue to receive dignified treatment. Therefore, employees should not be treated as passive objects of algorithmic decisions but as active organizational members who evaluate the legitimacy of technology use in the workplace. This perspective is essential because HRM decisions are closely related to employee welfare, career development, and perceptions of organizational justice.

The practical recommendations derived from this study should be understood as theory-informed governance implications rather than as separate effects directly tested by the survey. The empirical model measured Perceived AI Fairness, Algorithmic Trust, and Work Engagement; it did not separately estimate explainability, algorithmic auditing, appeal mechanisms, meaningful human review, preservation of dignity, or limits on algorithmic authority. Based on Human-Centered AI, organizational justice, and related AI-governance literature, organizations may nevertheless consider clearly communicating the role and limits of AI, providing channels to question or appeal consequential decisions, retaining meaningful human review for high-impact HR decisions, and periodically examining systems for potential bias or inconsistent treatment. These recommendations extend the empirical findings through established theory and should not be presented as interventions proven effective by this survey.

The pooled analysis should not be interpreted as evidence that all AI-HRM applications operate identically. Fairness judgments may differ across recruitment, performance appraisal, training, promotion, prediction, recommendation, and other HR uses because these applications vary in stakes, explainability, surveillance intensity, and opportunities for employee voice or appeal. Industry, organizational size, digital maturity, employee position, and exposure frequency may also condition the observed relationships. In addition, if several respondents came from the same organization, within-organization dependence could violate the assumption that all observations are fully independent. Because the retained data do not provide organization identifiers or adequate subgroup counts, clustering and multigroup robustness tests cannot be reported. Future data collection should record these characteristics prospectively, assess intraclass dependence where relevant, and establish measurement invariance before subgroup comparisons.

This study has several limitations. First, the cross-sectional design prevents strong causal conclusions and leaves reverse causality and omitted-variable endogeneity unresolved. Second, all constructs were self-reported at one time point, creating potential common-method and social-desirability bias; a full-collinearity or marker-variable assessment requires the original dataset. Third, the sample of 100 met the stated power criterion but remains limited for subgroup analysis and broad generalization. Fourth, the retained records do not provide recoverable counts by AI-HRM application, organization, respondents per organization, industry, firm size, digital maturity, employee position, or exposure frequency, so potential heterogeneity and non-independence cannot be evaluated. Fifth, Perceived AI Fairness was estimated as one reflective construct even though its seven indicators span procedural consistency, non-discrimination, outcome fairness, data relevance, clarity, voice, and

correctability; this limitation is especially relevant to H1, H2, and H4 and warrants comparison with a multidimensional or higher-order model. Finally, the available SmartPLS summary does not contain  $\rho_A$ , bootstrapped HTMT confidence intervals, outer and inner VIF,  $f^2$  by path,  $Q^2/PLSpredict$ , SRMR, or bias-corrected bootstrap confidence intervals. These diagnostics require a rerun of the original project before complete methodological compliance can be claimed. Future studies should use longitudinal or experimental designs, objective system characteristics, clustered or multilevel sampling where needed, and alternative construct specifications.

#### **4. Conclusion**

This study concludes that Perceived AI Fairness, Algorithmic Trust, and Work Engagement are positively associated in the eligible sample of employees directly exposed to AI-enabled HR processes. The positive direct association and statistically significant indirect pathway are consistent with a complementary mediation pattern, not causal proof that Algorithmic Trust transmits an effect of fairness onto engagement. The contribution lies in integrating organizational justice with the Job Demands–Resources perspective to interpret how employees evaluate AI-HRM and how those evaluations correspond with trust and engagement. Because Perceived AI Fairness was modeled as a single reflective construct despite conceptually heterogeneous domains, conclusions regarding H1, H2, and H4 should be interpreted at the level of global fairness perception and validated against higher-order alternatives in future research. Recommendations concerning explainability, algorithmic audits, appeal mechanisms, meaningful human review, dignity, and limits on algorithmic authority are theory-informed governance implications rather than constructs directly tested in this survey. Generalization remains conditional on AI application, organizational context, workforce composition, and exposure characteristics, and complete PLS-SEM diagnostic reporting still requires the original SmartPLS project.

#### **References**

- Acikgoz, Y., Davison, H. K., Compagnone, M., & Laske, M. (2020). Justice perceptions of artificial intelligence in selection. *International Journal of Selection and Assessment*, 28(4), 399–416. <https://doi.org/10.1111/ijsa.12306>
- Araujo, T., Helberger, N., Kruijmeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35, 611–623. <https://doi.org/10.1007/s00146-019-00931-w>
- Borst, R. T., Kruyen, P. M., Lako, C. J., & de Vries, M. S. (2020). The attitudinal, behavioral, and performance outcomes of work engagement: A comparative meta-analysis across the public, semipublic, and private sector. *Review of Public Personnel Administration*, 40(4), 613–640. <https://doi.org/10.1177/0734371X19840399>
- Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R., Boselie, P., Cooke, F. L., Decker, S., DeNisi, A., Dey, P. K., Guest, D., Knoblich, A. J., Malik, A., Paauwe, J., Papagiannidis, S., Patel, C., Pereira, V., Ren, S., Rogelberg, S., Saunders, M. N. K., Tung, R. L., & Varma, A. (2022). Artificial intelligence—Challenges and opportunities for international HRM: A review and research agenda. *The International Journal of Human Resource Management*, 33(6), 1065–1097. <https://doi.org/10.1080/09585192.2022.2035161>

- Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, 39(9), 1727–1739. <https://doi.org/10.1080/10447318.2022.2050543>
- Decuyper, A., & Schaufeli, W. (2020). Leadership and work engagement: Exploring explanatory mechanisms. *German Journal of Human Resource Management*, 34(1), 69–95. <https://doi.org/10.1177/2397002219892197>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. <https://doi.org/10.1145/3442188.3445923>
- Jarrah, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E., & Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society*, 8(2). <https://doi.org/10.1177/20539517211020332>
- 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>
- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13, 795–848. <https://doi.org/10.1007/s40685-020-00134-w>
- Kwon, K., & Kim, T. (2020). An integrative literature review of employee engagement and innovative behavior: Revisiting the JD-R model. *Human Resource Management Review*, 30(2), 100704. <https://doi.org/10.1016/j.hrmr.2019.100704>
- Langer, M., & Landers, R. N. (2021). The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, 123, 106878. <https://doi.org/10.1016/j.chb.2021.106878>
- Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, 160, 149–167. <https://doi.org/10.1016/j.obhdp.2020.03.008>
- Parker, S. K., & Grote, G. (2022). Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. *Applied Psychology*, 71(4), 1171–1204. <https://doi.org/10.1111/apps.12241>
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human–Computer Studies*, 146, 102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>

- Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review. *Big Data & Society*, 9(2). <https://doi.org/10.1177/20539517221115189>
- Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management*, 33(6), 1237–1266. <https://doi.org/10.1080/09585192.2020.1871398>
- Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Information Systems Frontiers*, 24(3), 857–875. <https://doi.org/10.1007/s10796-021-10223-8>
- Hair, J., & Alamer, A. (2022). Partial Least Squares Structural Equation Modeling (PLS-SEM) in second language and education research: Guidelines using an applied example. *Research Methods in Applied Linguistics*, 1(3), 100027. <https://doi.org/10.1016/j.rmal.2022.100027>
- Heveri, M., Novak, L., Poláčková Šolcová, I., & Tavel, P. (2025). Validation of the Utrecht work engagement scale (UWES-9) in the Czech Republic. *Scientific Reports*, 15, 42767. <https://doi.org/10.1038/s41598-025-26907-z>
- Köchling, A., Wehner, M. C., & Ruhle, S. A. (2025). This (AI)n't fair? Employee reactions to artificial intelligence (AI) in career development systems. *Review of Managerial Science*, 19(4), 1195–1228. <https://doi.org/10.1007/s11846-024-00789-3>
- Meijerink, J., & Bondarouk, T. (2023). The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. *Human Resource Management Review*, 33(1), 100876. <https://doi.org/10.1016/j.hrmr.2021.100876>
- Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S., & De Cremer, D. (2024). Fairness perceptions of artificial intelligence: A review and path forward. *International Journal of Human–Computer Interaction*, 40(1), 4–23. <https://doi.org/10.1080/10447318.2023.2210890>