AI Literacy and Ethical Voice in Algorithmic Hiring: The Mediating Role of Psychological Safety and Moral Efficacy

Muhammad Farhan Khan

Project Management Implementation Unit (PMIU), Health Department, Khyber Pakhtunkhwa, Pakistan

Abstract: Responsible human resource management (HRM) relies on human oversight, which works only when people are willing to question problematic algorithmic outputs. Based on social cognitive theory and psychological safety theory, this study investigates whether there is a direct link between AI literacy and the ethical voice in algorithms for hiring, as well as whether there is an indirect link through psychological safety and moral efficacy. Data of the survey in the form of information was collected from 390 employees who were engaged in recruitment and selection process in four multinational companies in Pakistan (purposive sampling; 470 questionnaires were distributed). The regression and bootstrap analyses revealed that the variables of psychological safety (B = 0.31) and moral efficacy (B = 0.37) had a positive relationship with AI literacy, while the direct path from ethical voice to AI literacy showed a stronger effect (total effect B = 0.44, $p < .001$). Psychological safety and moral efficacy both predicted ethical voice, and there was an indirect relationship between psychological safety and moral efficacy. Indirect effects made up approximately 42% of the overall effect, and there were significant indirect paths via psychological safety, via moral efficacy, and sequentially via both psychological safety and moral efficacy, in addition to a significant direct path (complementary partial mediation). AI-literacy training in combination with a safe reporting climate and high moral efficacy might therefore be most likely to help supervise ethical use. Data are cross-sectional in nature and self-reported, so the results are interpreted in terms of associations.

Keywords: AI Literacy, Ethical Voice, Psychological Safety, Moral Efficacy, Algorithmic Hiring, Human Oversight, Pakistan

Email: farhanphd1@gmail.com

1. Introduction

Today, artificial intelligence (AI) is common in human resource management (HRM) in recruitment and selection processes Hunkenschroer & Luetge [1]; Kim [2]; Tambe [3]. It is not just advanced economies – 43% of people questioned at OICCI in Pakistan report that they are already using generative AI, while 81% anticipate that AI will assume vital roles in their enterprises (Aazim, 2026). Algorithmic hiring is likely to become a tool for everyday managerial practice in the case of multinational subsidiaries that tend to employ formalized HR practices more intensely than domestic firms Ahmad [4].

The ethical risk these efficiency gains present. The biases of collected data and design can be mimicked when using algorithmic screening services Kelan [5], Naoum [6], and algorithms can be seen as less valuable, dignified, and fair than human screening Khan & Khan [7]; Bankins [8]. Responsible AI-supported HRM thus prioritizes human monitoring at the centre of the practice

Bankins [9]; Rodgers [10]); however, if there is no challenger to the system, then the human monitoring is not effective. Questioning, for instance, a black-box score, or disclosing the adverse impact is, for example, a form of ethical voice, defined as discretionary upward communication of ethical risks, unfairness, or questionable practices Liang [11]; Morrison [12]. Voice is interpersonally risky – 93% would voice, but only 50% do Liang [11]. That means it is an open question as to what allows hiring professionals to speak up.

AI literacy, the abilities necessary to critically assess, communicate, and collaborate with AI Long & Magerko, [13], is a natural candidate, but evidence of its protective value is mixed Khan & Khan, [7]; Fernandes [14]; Horowitz & Kahn, [15]; Klingbeil [16]: knowledge seems necessary yet not sufficient, and making it a behavior is a product of psychological and social factors. We focus on two theories: psychological safety, as a reduction in the interpersonal cost of speaking up Edmondson, [17]; and moral efficacy, or an individual's self-confidence in their ability to mobilize motivation and resources to take action on moral judgments Bandura, [18]; Hannah & Avolio, [19]. Both are consistently associated with ethical voice Han [20]; Lee [21]; Liao [22]; Sert & Mert [23], and a qualitative study suggests that psychological safety is part of how AI literacy enables challenging decisions made by AI Vanisri [24] although this has not been tested in hiring.

Studies on adoption and perceived impacts of AI in HRM (Khan & Khan, [25]; Irfan, M., & Ilyas, M.; [26]; Ahmed [27] have focused on adoption and expected impacts of AI in HRM, but not conditions for ethical behavior, and evidence on human–machine interaction in AI-based hiring is rare Kelan, [5]. We are aware of no research that examines the psychological safety and confidence that AI literacy provides to hiring professionals in raising moral concerns, especially in multinational subsidiaries where hierarchical influences might also increase the interpersonal costs of contesting a system that is endorsed by the top of the hierarchy. Accordingly, we examine a model in which the relationships between AI literacy and ethical voice are both direct and indirect via psychological safety and moral efficacy, with survey responses obtained from 320 employees of four multinationals in Pakistan (PepsiCo, Coca-Cola, Nestlé, and Unilever).

The study has three contributions. It does so by taking the ethics research of AI-HRM from the level of governance principles to one of actual outcomes, involving ethical voice and challenging the assumption of human oversight that underpins governance principles. Second, it combines three constructs: a personal capability (AI literacy), a perception of social context (psychological

safety), and a self-regulatory belief (moral efficacy) into one serial mediation model, which involves an under-studied relationship between safety and efficacy. Third, it offers new support for algorithmic hiring by multinational subsidiaries in an emerging economy.

Literature review and Hypothesis development.Literature review and Hypothesis development.

2. Literature Review and Hypothesis Development

2.1 Algorithmic hiring and the challenge of ethical oversight

The studies on 'AI + HRM' have shifted from putting forth claims of efficiency to concerns about fairness, dignity, accountability, and legitimacy Hunkenschroer & Luetge, [1]; Qadri, [28]. Human oversight is recommended Bankins, [9]; Rodgers [10], yet it can be undermined by flawed data and/or design, which can recreate the discrimination mechanism Köchling & Wehner, [29]. In one study, recruiters were more influenced by algorithmic than human advice, despite a lower level of trust in the former; overreliance was experienced by experts and did not change after training; and inaccurate advice hurt decision-making more than accurate advice did Lacroux & Martin Lacroux, [30]; recruiters with volume targets may delegate ethical decision-making to higher-level management Dutta & Naveen, [31]; and overreliance is found among experts and is not reduced by training Parasuraman & Manzey, [32]; Romeo & Conti, [33].

2.2 AI literacy: concept, measurement, and a double-edged role

AI Literacy is typically interpreted as the skills needed for people to critically assess, communicate and collaborate with, and use AI in their professional lives Long & Magerko, [13] and, in the organisational context, as the knowledge of a worker about how to use and perceive the influence of AI systems Çetindamar [34]. Validated, mostly self-report, scales are consistent in their concepts of technical knowledge, application, critical assessment, and ethics (Lintner [35]; Wang [36]. Evaluation and ethics are most important for HR professionals, as they can identify if the ranking is biased based on history, in which case they will decide to overrule it. However, the protective effects are unclear: There is potential for people to take advice from AI that contradicts the cue information. Klingbeil [16] notes that moderate knowledge may be the most vulnerable to automation bias (Horowitz & Kahn [15] and more knowledge of AI may lead to less accurate self-assessment Fernandes [14]. Literacy alone is unlikely to be sufficient; the changes to behaviour are reliant on psychological and social factors.

2.3 Ethical voice, psychological safety, and moral efficacy

Voice is when employees bring ideas, concerns, and opinions upward (discretionary), and silence is when they withhold ideas, concerns, and opinions Morrison [37]. There are two kinds: promotive (new ideas) and prohibitive (concerns about harmful practices) Chamberlin [38]; Liang [39]. Ethical voice involves an ethical stake, including ethical risks of voicing, violations of fairness, and questionable practices Lee [21]; Liang [11]; it is discretionary, interpersonally risky, and a moral choice to speak up in any way, which makes it both a communicative and a moral decision Knoll [40].

Two mechanisms are particularly pertinent. Psychological safety is a shared sense of the team's safety for taking interpersonal risks Edmondson [17] and was associated with voice, reducing the perceived costs of voicing (Edmondson & Lei [41]; Frazier [42]; Tedone [43]). It was also more strongly related to voice over acquiescent silence than individual courage was among port workers Sert & Mert [23] and was linked to ethical leadership Burhan [44].

With moral ownership and courage, moral efficacy becomes moral potency Hannah & Avolio [19] which is the belief one has in their own ability to muster the motivation and resources to act on moral decisions. It connects moral potency, ethical voice, and ethically focused proactive work behavior Gok [45], and ethical voice, psychological safety, and peer reporting (Sumanth [46], as well as ethical leadership and ethical voice Lee [21] and responsible leadership and ethical voice Han [20]; Irfan, M., & Khan [47], respectively.

3. Theoretical lens and hypotheses

Social cognitive theory suggests that humans are agents of their actions, guided by efficacy beliefs and social structures Bandura, [18]; psychological safety theory suggests that the influence between efficacy beliefs and actions is the interpersonal-risk calculus Edmondson, [17]; Edmondson & Lei, [41]. AI literacy is viewed as a person-level capability, moral efficacy is considered as the self-regulatory construct between capability and behavior, and psychological safety is understood as the perceived social context that impacts efficacy.

3.1 AI literacy and psychological safety

Literate professionals can raise concerns in the language of the tool (data, features, thresholds), minimising the interpersonal risk of coming across as in opposition to technology (Edmondson & Lei [41] and literacy reduces the fear of incompetence and uncertainty that I feel is created when the system is opaque. Adjacent evidence: Psychological safety is related to challenging

decisions made by AI Vanisri [24]; psychological safety impacts AI utilization and employee outcomes (Kim [48]; and being aware of AI without the competencies can create an insecure and silent experience Cheng [49]. *H1: AI literacy is positively related to psychological safety.*

3.2 AI literacy and moral efficacy

Social cognitive theory identifies knowledge and mastery as sources of efficacy Bandura [18]. Literate HR professionals have the evaluative and ethical competencies to identify and judge problematic outputs Wang [36], raising their confidence in mobilizing moral action Hannah & Avolio [19], whereas those who sense a problem but lack the know-how to address it may feel uneasy without knowing how to respond Passlack [50]. *H2: AI literacy is positively related to moral efficacy.*

3.3 Psychological safety and moral efficacy

Safe climates allow people to discuss emerging issues and make mistakes without repercussion (Edmondson, 1999) and thus acquire the enactive and vicarious experiences that are the basis for efficacy development (Bandura, 2001). Moral potency research largely supports this as well: psychological safety and moral potency combine to positively predict peer reporting (Sumanth et al., 2024), and ethical psychological climates predict moral potency (Gok et al., 2022). There is little direct evidence to support or refute the safety–efficacy link, and it is only a theory that needs to be tested. *H3: Psychological safety is positively related to moral efficacy.*

3.4 Psychological safety, moral efficacy, and ethical voice

Psychological safety is one of the most common indicators of voice and decreased silence, including ethical voice (Burhan et al., 2023); it decreases the cost of speaking up. The effects of antecedents on ethical voice have been mediated by moral efficacy, as it provides the confidence to speak because ethical contexts are translated into ethical voice (Han et al., 2024; Lee et al., 2017; Liao et al., 2022). *H4: Psychological safety is positively related to ethical voice. H5: Moral efficacy is positively related to ethical voice.*

3.5 AI literacy, ethical voice, and mediation

Literacy should have a direct connection to ethical voice, as it provides the information and knowledge necessary to identify a problem that needs to be addressed. But there is an overdependence even with literacy and training (Klingbeil et al., 2024; Parasuraman & Manzey, 2010), and it is highly probable that literacy alone won't yield voice. Thus, we would expect it to work in two ways: first, that literate professionals feel better about challenging algorithmic

outputs (H1), and then that safety contributes to their sense of competence to do so (H3), which in turn supports ethical voice (H5). *H6: AI literacy is positively related to ethical voice.*

H7a: Psychological safety mediates the relationship between AI literacy and ethical voice.

H7b: Moral efficacy mediates the relationship between AI literacy and ethical voice.

H7c: AI literacy influences ethical voice sequentially through psychological safety and then moral efficacy.

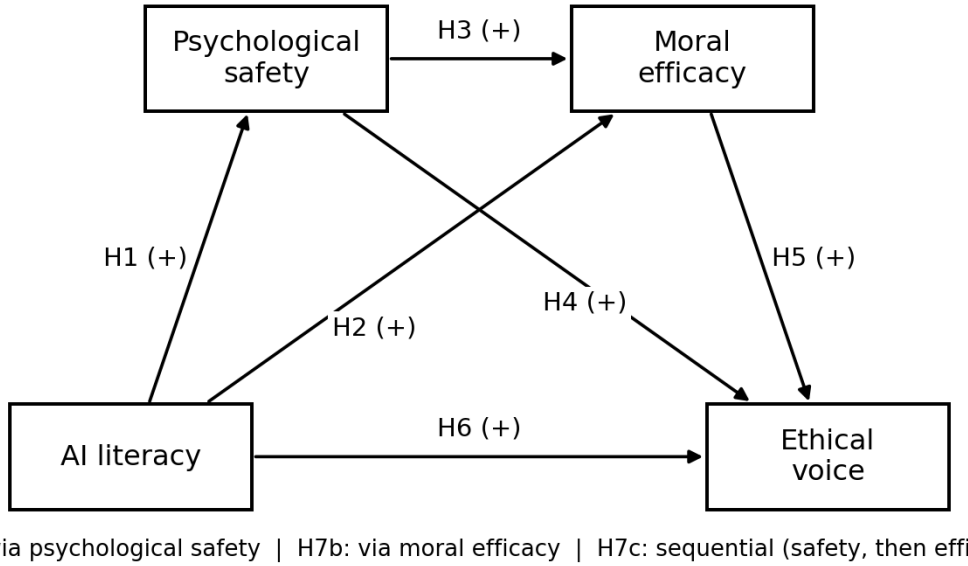


Figure 1: Conceptual Model

4. Method

4.1 Research design, sample, and procedure

To test the model, a cross-sectional survey was conducted in four multinational companies in Pakistan (PepsiCo, Coca-Cola, Nestlé, and Unilever) via purposive sampling with the requirement of relevant exposure with employees Etikan [51]. The participants included recruiters, HR generalists/business partners, HR managers/directors, and line managers who hire; a cover note provided confidentiality and anonymity. In total, 390 questionnaires were usable (83.0%), which is more than the minimum of 10 questionnaires for a factor analysis (FA) for each questionnaire item (17 questionnaires for 23 items; Hair [52].

Recruiters/talent acquisition specialists (25.4%), HR generalists/business partners (23.6%), hiring line managers (22.8%), HR managers/directors (19.5%), and others (8.7%) responded,

with 82.8% having at least two years of HR or recruitment experience and 69.2% using AI tools for screening or selection daily or weekly.

4.2 Measures

Items were adapted for rating all constructs on 5-point scales (1 = strongly disagree, 5 = strongly agree) (full wording in Appendix A as supplemental material). The sample was described by single items on role, experience, and frequency of use of AI tools.

4.3 AI literacy

A self-perceived ability, operational knowledge, and critical evaluation skills were assessed using six items of the Perceived Artificial Intelligence Literacy Questionnaire (PAILQ-6; Grassini, [53]; $\alpha = .81$).

4.4 Psychological safety

Edmondson's (1999) 7-point scale was adapted to fit the context of AI-supported hiring (e.g., "It is safe to take a risk by questioning/overriding an AI-driven hiring recommendation in this organization"). Items 3 and 5 were reverse-coded (6 - original score); $\alpha = .85$.

4.5 Moral efficacy

Five items adapted from Hannah [54] and Lee et al. (2017) assessed confidence in organizing and taking action on moral issues ($\alpha = .84$).

4.6 Ethical voice

The 5 items utilized in this study to assess proactive reporting of ethical concerns and potential misconduct in AI practices ($\alpha = .83$) were adapted from Lee et al. (2017) and Huang and Paterson [55].

4.7 Data analysis

Reliability, construct validity, and common method bias were assessed with Cronbach's α ($\geq .70$; Nunnally & Bernstein [56], principal component analysis (PCA) (varimax rotation, eigenvalues > 1) with the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test Kaiser [57], and Harman's single-factor test Podsakoff [58]. Ordinary least squares regression was used to test the hypotheses H1-H6, and serial mediation was tested with 5,000 bootstrap re-sample with 95% confidence interval (CI) Preacher & Hayes, [59]; mediation type was based on Zhao [60], indirect effect was significant when CI did not contain zero. Scale means (after reverse coding) were used as scores, and unstandardized (B) and standardized (β) coefficients are provided. IBM SPSS 27 was used for analyses.

5. Results

5.1 Measurement quality and common method bias

Reliability and validity. All scales had Cronbach's α above .70, ranging from .81 to .85 (Table 1). Overall KMO was .91, and Bartlett's test was significant, $\chi^2(253) = 3{,}253.21$, $p < .001$. Four components of varimax-rotated PCA were obtained that explained 56.10% of variance and contained six AI literacy, seven psychological safety, five moral efficacy, and five ethical voice items. For the most part, items were loaded on their intended component (.66–.77) without being cross-loaded on any components > .26, and reverse-coded items PS3 and PS5 were loaded on psychological safety, with scores of .69 and .77, respectively.

Discriminant validity and collinearity. Constructs had moderate to low correlation with each other (0.31 – 0.48; Table 1), which is far from .85 (empirical indistinguishability; Kline, [61]. The variance inflation factors were not greater than 1.3, which provides an indication that there was no collinearity.

Common method bias. We ran the test by Harman (1976), and the first unrotated component accounted for 29.67% of the variance, which is less than the 50% cut-off value, since the test might not be sensitive enough to the extent of common method variance Podsakoff [58].

5.2 Descriptive statistics and correlations

Descriptive statistics and correlations are presented in Table 1. Means were close to the scale midpoint (2.93-3.05), and all correlations were positive and significant ($p < .001$), with ethical voice correlating with moral efficacy ($r = .48$), AI literacy ($r = .41$), and psychological safety ($r = .40$).

Table 1: Descriptive Statistics, Reliability, and Correlations among Study Variables

Variable	M	SD	α	1	2	3	4
1. AI literacy	3.05	0.89	.81	—			
2. Psychological safety	3.00	0.90	.85	.31	—		
3. Moral efficacy	2.99	0.99	.84	.33	.35	—	
4. Ethical voice	2.93	0.95	.83	.41	.40	.48	—

Note. $N = 390$. Scores range from 1 to 5; psychological safety was computed after reverse-coding items 3 and 5. α = Cronbach's alpha. All correlations are significant at $p < .001$.

5.3 Tests of the direct relationships (H1–H6)

Table 2 reports the regression results, and Figure 1 shows the estimated model.

H1–H3. AI literacy was positively related to psychological safety (B = 0.309, β = .305, p < .001) and moral efficacy (B = 0.369, β = .332, p < .001), supporting H1 and H2. With AI literacy controlled, psychological safety was positively related to moral efficacy (B = 0.296, β = .270, p < .001), supporting H3.

H4–H6. The three predictors together explained 34.2% of the variance in ethical voice, F(3, 386) = 66.85, p < .001. Psychological safety (B = 0.222, β = .210) and moral efficacy (B = 0.320, β = .332) each predicted ethical voice (both p < .001), supporting H4 and H5, and AI literacy retained a significant direct relationship (B = 0.255, β = .238, p < .001). The overall effect of AI literacy on ethical voice is also positive (B = 0.442, β = .412, p < .001), which supports H6.

Table 2: Regression Results for the Hypothesized Relationships

Outcome/predictor	B	SE	β	t	p	95% CI	R^2
Model 1: Psychological safety (H1)							.093
AI literacy	0.309	0.049	.305	6.31	< .001	[0.213, 0.405]	
Model 2: Moral efficacy (H2)							.110
AI literacy	0.369	0.053	.332	6.93	< .001	[0.265, 0.474]	
Model 3: Moral efficacy (H3)							.176
AI literacy	0.278	0.054	.249	5.15	< .001	[0.172, 0.384]	
Psychological safety	0.296	0.053	.270	5.57	< .001	[0.192, 0.401]	
Model 4: Ethical voice (H4, H5)							.342
AI literacy	0.255	0.048	.238	5.30	< .001	[0.161, 0.350]	
Psychological safety	0.222	0.048	.210	4.66	< .001	[0.128, 0.316]	
Moral efficacy	0.320	0.044	.332	7.30	< .001	[0.234, 0.406]	
Model 5: Ethical voice, total effect (H6)							.170
AI literacy	0.442	0.050	.412	8.91	< .001	[0.344, 0.539]	

Note. N = 390. B = unstandardized coefficient; β = standardized coefficient; CI = confidence interval. Intercepts are omitted. All models were significant at $p < .001$.

5.4 Mediation analysis (H7a–H7c)

Bootstrap results are shown in Table 3. Indirect effects via psychological safety (B = 0.069; H7a) or moral efficacy (B = 0.089; H7b) or both (B = 0.029; H7c) were all significant (bootstrap 95% CIs did not contain zero). Its total effect was 0.187, with the direct effect significant and the

same in sign as the total effect (B = 0.255); thus, the partial mediation was complementary, and the indirect paths accounted for 42.3% of the total effect (B = 0.442). Even though we found that the moral efficacy path was larger than the psychological safety path, this was not formally tested. The results of all hypotheses were accepted.

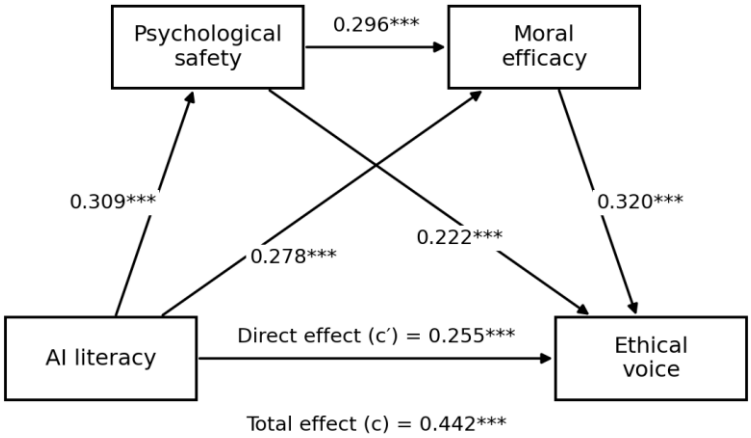


Figure 1: ***Estimated Serial Mediation Model***

Note. N = 390. Unstandardized coefficients. ***p < .001. The path from AI literacy to moral efficacy is estimated controlling for psychological safety, and the direct path to ethical voice is estimated controlling for both mediators.

Table 3: ***Bootstrap Estimates of Direct, Indirect, and Total Effects of AI Literacy on Ethical Voice***

Effect	B	Boot SE	95% bootstrap CI	% of total
AI literacy → Psychological safety → Ethical voice (a_1b_1)	0.069	0.018	[0.036, 0.109]	15.6
AI literacy → Moral efficacy → Ethical voice (a_2b_2)	0.089	0.021	[0.051, 0.133]	20.1
AI literacy → Psychological safety → Moral efficacy → Ethical voice ($a_1d_{21}b_2$)	0.029	0.008	[0.016, 0.046]	6.6
Total indirect effect	0.187	0.029	[0.133, 0.247]	42.3
Direct effect (c′)	0.255	0.046	[0.163, 0.347]	57.7
Total effect (c)	0.442	0.046	[0.349, 0.532]	100.0

Note. N = 390. Unstandardized coefficients; 5,000 bootstrap resamples. An effect is significant when its confidence interval excludes zero (all effects shown are significant). The percentage of the total effect was computed from the point estimates.

6. Discussion

6.1 Summary and theoretical implications

All hypotheses were supported. The predictors accounted for 34.2% of the variance in ethical voice, with approximately 42% of the total effect of AI literacy being indirect.

Firstly, the conclusions move the focus in AI-HRM ethics away from governance principles towards a behavioural outcome. Although frameworks tend to place human oversight as their heart and soul (Bankins, 2021; Rodgers et al., 2022), algorithmic suggestions can still influence recruiters despite lower levels of trust (Lacroux & Martin Lacroux, 2026). These findings led us to believe that the decision to hire professionals is systematically different in terms of expressing concerns, and that this is linked to their self-confidence and understanding of the surrounding environment. Although self-perceived literacy is positively correlated with voice, this does not invalidate the findings of Fernandes et al. (2026) and Klingbeil et al. (2024), who focused on reliance and metacognition, respectively; perhaps self-perceived literacy could protect by raising concern without enhancing individual judgments. A curvilinear relationship was not modeled, where moderate knowledge might prove to be most risky (Horowitz & Kahn, 2023).

Second, combining a self-efficacy belief, a perception of social context, and a personal capability reveals the connection between literacy and voice: all three of the indirect pathways were statistically significant, and psychological safety and moral efficacy were uniquely related to ethical voice. The safety–efficacy link (H3) resonates with a social cognitive perspective that views efficacy as emerging from enactive or vicarious experience (Bandura, 2001) and is easier when concerns can be openly shared without repercussions (Edmondson, 1999), and contributes to evidence that safety and moral potency combined facilitate ethical action (Gok et al., 2022; Sumanth et al., 2024) in the context of AI-supported hiring. While we did not test this difference, moral efficacy had the larger standardized coefficient ($\beta = .33$ vs. .21), consistent with its role as the proximal belief that converts ethical contexts into action (Han et al., 2024; Lee et al., 2017; Liao et al., 2022), though we did not test this difference as it was only between the port terminal and operational safety.

Third, partial mediation indicates that social and motivational conditions act as a facilitator of the relation between literacy and voice. The indirect routes combined (42.3%) were smaller than the direct route, indicating two levels: literacy to the perception and efficacy to speak up, and literacy to get the knowledge to “discover” and articulate a problem to speak up about. Direct

effects can also be the consequence of an omitted mechanism, e.g., taking responsibility for recruitment results or common method variance. The results align with qualitative studies about psychological safety as one of the paths through which literacy can enable individuals to challenge AI-made decisions (Vanisri et al., 2026); however, we tested mediation, not moderation, so we cannot conclude that literacy enables individuals to challenge the AI-based decisions only when they feel safe. Lastly, capability measures close to the scale midpoint indicate that ethical oversight is minimal, although present, and we did not assess whether hierarchical norms influenced capabilities.

6.2 Practical implications

These suggestions must be tested in intervention studies, as the data are cross-sectional.

Provide AI-literacy training along with fostering safe reporting environments and moral competency. Programs should include scenarios in which a skewed shortlist, opaque score, or adverse impact concern is flagged, and mastery experiences are provided to build moral efficacy (Bandura, 2001).

Explicitly make the algorithm safe to be challenged. Two ways to communicate that it's okay to disagree with AI recommendations are through documented override and escalation processes (including to the vendor) and through visible follow-up actions taken with concerns raised, with no blame attributed to the users for any errors if they choose to disagree with AI suggestions.

Treat the 3 conditions as a diagnostic; don't rely on human review alone. A short pulse survey of AI literacy, psychological safety, and moral efficacy could help guide the design of a responsible-AI audit process (a unit that is high in literacy but low in safety needs a climate action/process rather than more training); governance policies should include what to do as part of the training, the climate, and reporting process that makes human review effective.

7. Limitations and future research

This study has limitations. Firstly, results of mediation are patterns consistent with the proposed sequence, and whether or not they are evidence is debatable (e.g., successful speaking up could contribute to feelings of safety and efficacy); effects could be reciprocal, and longitudinal or experimental designs are required. Second, all variables were self-reported at the same time point from the same respondents, and Harman's test is insensitive (Podsakoff et al., 2003); social desirability may make it easier to express voice, and perceived proficiency may not correspond to actual proficiency (Fernandes et al., 2026). Objective measures of literacy, supervisor and/or

peer-recorded measures of voice and behaviour, and latent-variable models should be used in future research.

Third, the use of purposive sampling in four multinationals is subject to limitations in terms of generalizability, and the respondents' firms were not recorded to allow modelling of clustering; multilevel designs and replication in domestic firms, other sectors and other cultures are required. Fourth, the scales used were adapted; validity was tested using PCA instead of CFA; psychological safety was measured at the individual level, not at the team level (cf. Edmondson, 1999), and role, experience, and AI-use frequency were not controlled. Fifth, we looked neither at whether voice was heard, acted upon, nor whether it made hires more fair, but at whether people voiced concerns. Future research is warranted using the CFA, using a team-referent measure, and using covariates, tracing voice episodes to outcomes, examining boundary conditions (e.g., ethical leadership, power distance, algorithmic explainability), and testing curvilinear effects of literacy (Horowitz & Kahn, 2023) and psychological safety (Vanisri et al., 2026).

8. Conclusion

Human oversight of algorithmic hiring requires capable and confident individuals to question it. The relationship between AI-literacy and ethical voice was mediated by psychological safety and moral efficacy, with AI-literacy being directly linked to ethical voice and psychological safety being directly linked to moral efficacy among employees at four multinationals in Pakistan, suggesting that AI-literacy programs will be most effective in conjunction with establishing safe reporting climates and promoting moral efficacy. Based on this cross-sectional, self-reported, purposive-sample evidence, future research would involve testing these pathways across time with multi-source data with other samples and assessing if the voice they enable leads to increased fairness in hiring.

References

[1]. Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics, 178*, 977–1007. https://doi.org/10.1007/s10551-022-05049-6

[2]. Kim, B.-J., Kim, M.-J., & Lee, J. (2025). The dark side of artificial intelligence adoption: Linking artificial intelligence adoption to employee depression via psychological safety and ethical leadership. *Humanities and Social Sciences Communications, 12*. https://doi.org/10.1057/s41599-025-05040-2

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

[4]. Ahmad, M., Allen, M. M. C., Raziq, M. M., & Rehman, W. U. (2019). Converging HRM practices? A comparison of high performance work system practices in MNC subsidiaries and domestic firms in Pakistan. *Employee Relations, 41*(5), 931–948. https://doi.org/10.1108/ER-01-2018-0021

[5]. Kelan, E. K. (2023). Algorithmic inclusion: Shaping the predictive algorithms of artificial intelligence in hiring. *Human Resource Management Journal, 34*(3), 694–707. https://doi.org/10.1111/1748-8583.12511

[6]. Naoum, R. F., Szakadáti, T., & Balogh, G. (2026). Artificial intelligence (AI) in human resource management (HRM): A systematic review of its dual impact on diversity, equity, and inclusion (DEI). *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00580-y

[7]. Khan, M. F., & Khan, I. U. (2026a). AI-enabled HRM practices and employee innovative work behavior: The mediating role of psychological empowerment. Journal of Social Research Development, 7(4), 1-13.

[8]. 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*, 857–875. https://doi.org/10.1007/s10796-021-10223-8

[9]. Bankins, S. (2021). The ethical use of artificial intelligence in human resource management: A decision-making framework. *Ethics and Information Technology, 23*, 841–854. https://doi.org/10.1007/s10676-021-09619-6

[10]. Rodgers, W., Murray, J. M., Stefanidis, A., Degbey, W. Y., & Tarba, S. Y. (2022). An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review*. https://doi.org/10.1016/j.hrmr.2022.100925

[11]. Liang, S. (2026). Speaking up, acting out: The moral licensing dynamics linking ethical voice to unethical behavior. *Applied Psychology, 75*(3). https://doi.org/10.1111/apps.70105

[12]. Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals, 5*(1), 373–412. https://doi.org/10.1080/19416520.2011.574506

[13]. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3313831.3376727

[14]. Fernandes, D., Villa, S., Nicholls, S., Haavisto, O., Buschek, D., Schmidt, A., Kosch, T., Shen, C., & Welsch, R. (2026). AI makes you smarter but none the wiser: The disconnect between performance and metacognition. *Computers in Human Behavior, 175*, Article 108779. https://doi.org/10.1016/j.chb.2025.108779

[15]. Horowitz, M. C., & Kahn, L. (2023). Bending the automation bias curve: A study of human and AI-based decision-making in national security contexts. *arXiv*, Article 2306.16507. https://doi.org/10.48550/arxiv.2306.16507

[16]. Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior, 160*, Article 108352. https://doi.org/10.1016/j.chb.2024.108352

[17]. Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383. https://doi.org/10.2307/2666999

[18]. Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

[19]. Hannah, S. T., & Avolio, B. J. (2010). Moral potency: Building the capacity for character-based leadership. *Consulting Psychology Journal: Practice and Research, 62*(4), 291–310. https://doi.org/10.1037/a0022283

[20]. Han, Z.-Y., Zhu, Z.-Y., Wang, D., & Wei, W. (2024). Responsible leadership and employee ethical voice: Mediating role of ethical efficacy and moderating role of moral identity. *Current Psychology, 43*, 29516–29527. https://doi.org/10.1007/s12144-024-06564-7

[21]. Lee, D., Choi, Y., Youn, S., & Chun, J. U. (2017). Ethical leadership and employee moral voice: The mediating role of moral efficacy and the moderating role of leader–follower value congruence. *Journal of Business Ethics, 141*, 47–57. https://doi.org/10.1007/s10551-015-2689-y

[22]. Liao, Z., Cheng, J.-L., & Chen, Q. (2022). Socially responsible human resource management and employee ethical voice: Roles of employee ethical self-efficacy and organizational identification. *Corporate Social Responsibility and Environmental Management*. https://doi.org/10.1002/csr.2236

[23]. Sert, Ö., & Mert, İ. S. (2026). Silent weak signals in port operations: Ethical leadership, psychological safety and reporting climate in safety-critical terminals. *Journal of Contingencies and Crisis Management, 34*(3). https://doi.org/10.1111/1468-5973.70207

[24]. Vanisri, M., Kondapaka, P., & Sharma, H. (2026). AI literacy and organisational resilience in AI-enabled work: How psychological safety enables voice and adaptive learning – A grounded theory study. *International Journal of Drug Delivery Technology*. https://doi.org/10.25258/ijddt.16.26s.114

[25]. Khan, M. F., & Khan, I. U. (2026b). Algorithmic transparency, perceived fairness, and employee trust in AI-driven HR decisions: A mediation model. Journal of Social Sciences Development, 5(3), 40-52.

[26]. Irfan, M., & Ilyas, M. (2026). AI-TECHNOLOGY TRUST AS MEDIATOR IN LINKING PERCEIVED CHANGES IN HR FUNCTIONS AND EMPLOYEE ADOPTION INTENTIONS. JOURNAL OF SOCIAL SCIENCES DEVELOPMENT, 5(3), 53-64.

[27]. Ahmed, B., Ramish, M. S., & Mutasam. (2024). Artificial intelligence and future HRM practices: A case of Pakistan business sector. *Journal of Development and Social Sciences, 5*(2), 395–405. https://doi.org/10.47205/jdss.2024(5-II)37

[28]. Qadri, U. (2026). AI-enabled HRM as an ethical system: A systematic review of fairness, accountability, governance and legitimacy. *International Journal of Ethics and Systems*. https://doi.org/10.1108/ijoes-04-2026-0308

[29]. 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*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

[30]. Lacroux, A., & Martin Lacroux, C. (2026). When recruiters meet algorithms: Advice taking and bias in AI-assisted recruitment. *International Journal of Selection and Assessment, 34*(4). https://doi.org/10.1111/ijsa.70085

[31]. Dutta, D., & Naveen, P. M. (2025). Transforming recruitment and selection practices in organizations through discriminative and generative AI adoption: A structuration lens. *Human Resource Management, 65*(1), 77–115. https://doi.org/10.1002/hrm.70018

[32]. Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

[33]. Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. *AI & SOCIETY, 41*, 259–278. https://doi.org/10.1007/s00146-025-02422-7

[34]. Çetindamar, D., Kitto, K., Wu, M.-J., Zhang, Y., Abedin, B., & Knight, S. (2022). Explicating AI literacy of employees at digital workplaces. *IEEE Transactions on Engineering Management*. https://doi.org/10.1109/tem.2021.3138503

[35]. Lintner, T. (2024). A systematic review of AI literacy scales. *npj Science of Learning, 9*. https://doi.org/10.1038/s41539-024-00264-4

[36]. Wang, B., Rau, P.-L. P., & Yuan, T. (2022). Measuring user competence in using artificial intelligence: Validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology, 42*(9), 1324–1337. https://doi.org/10.1080/0144929x.2022.2072768

[37]. Morrison, E. W. (2014). Employee voice and silence. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 173–197. https://doi.org/10.1146/annurev-orgpsych-031413-091328

[38]. Chamberlin, M., Newton, D. W., & LePine, J. A. (2017). A meta-analysis of voice and its promotive and prohibitive forms: Identification of key associations, distinctions, and future research directions. *Personnel Psychology, 70*(1), 11–71. https://doi.org/10.1111/peps.12185

[39]. Liang, J., Farh, C. I. C., & Farh, J.-L. (2012). Psychological antecedents of promotive and prohibitive voice: A two-wave examination. *Academy of Management Journal, 55*(1), 71–92. https://doi.org/10.5465/amj.2010.0176

[40]. Knoll, M., Fida, R., Marzocchi, I., Searle, R., Connelly, C., & Ronchetti, M. (2025). Quiet workaholics? The link between workaholism and employee silence and moral voice as explained by the social-cognitive theory of morality. *Journal of Organizational Behavior, 46*(5), 745–764. https://doi.org/10.1002/job.2867

[41]. Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

[42]. Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. *Personnel Psychology, 70*(1), 113–165. https://doi.org/10.1111/peps.12183

[43]. Tedone, A. M., Mesmer-Magnus, J., Viswesvaran, C., & Newkirk, E. (2025). See something, say something: A meta-analytic review of safety voice. *International Journal of Management Reviews, 28*(1). https://doi.org/10.1111/ijmr.12401

[44]. Burhan, Q., Khan, M. A., & Malik, M. F. (2023). Ethical leadership: A dual path model for fostering ethical voice through relational identification, psychological safety, organizational identification, and psychological ownership. *RAUSP Management Journal*. https://doi.org/10.1108/rausp-01-2023-0008

[45]. Gok, K., Babalola, M. T., Sumanth, J. J., Lakshman, C., Vo, L., Decoster, S., Bansal, A., & Coskun, A. (2022). Enhancing employees' duty orientation and moral potency: Dual mechanisms linking ethical psychological climate to ethically-focused proactive behaviors. *Journal of Organizational Behavior*. https://doi.org/10.1002/job.2667

[46]. Sumanth, J. J., Hannah, S. T., Herbst, K., & Thompson, R. L. (2024). Generating the moral agency to report peers' counterproductive work behavior in normal and extreme contexts: The generative roles of ethical leadership, moral potency, and psychological safety. *Journal of Business Ethics, 195*, 653–680. https://doi.org/10.1007/s10551-024-05679-y

[47]. Irfan, M., & Khan, I. U. (2024). The Impact Of Inclusive Leadership On Innovative Performance: Mediating Role Of Intellectual Stimulation. Journal Of Social Sciences Development, 3(4), 40-51.

[48]. Kim, S., Khoreva, V., & Vaiman, V. (2025). Strategic human resource management in the era of algorithmic technologies: Key insights and future research agenda. *Human Resource Management, 64*(2), 447–464. https://doi.org/10.1002/hrm.22268

[49]. Cheng, M.-T., Zhang, L., & Wang, H. (2025). The effect of artificial intelligence awareness on frontline service employees' silence: The roles of psychological contract breach and moral identity. *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/ijchm-07-2024-0968

[50]. Passlack, N., Hammerschmidt, T., & Posegga, O. (2026). With great power comes great responsibility: What shapes AI literacy for responsible interactions of knowledge workers with AI? *Information Systems Frontiers, 28*(1), 11–46. https://doi.org/10.1007/s10796-025-10648-5

[51]. Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4. https://doi.org/10.11648/j.ajtas.20160501.11

[52]. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

[53]. Grassini, S. (2024). A psychometric validation of the PAILQ-6: Perceived Artificial Intelligence Literacy Questionnaire. In *Proceedings of the 13th Nordic Conference on Human-Computer Interaction (NordiCHI 2024)*. ACM. https://doi.org/10.1145/3679318.3685359

[54]. Hannah, S. T., Avolio, B. J., & May, D. R. (2011). Moral maturation and moral conation: A capacity approach to explaining moral thought and action. *Academy of Management Review, 36*(4), 663–685. https://doi.org/10.5465/amr.2010.0128

[55]. Huang, L., & Paterson, T. A. (2017). Group ethical voice: Influence of ethical leadership and impact on ethical performance. *Journal of Management, 43*(4), 1157–1184. https://doi.org/10.1177/0149206314546195

[56]. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

[57]. Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

[58]. Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

[59]. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

[60]. Zhao, X., Lynch, J. G., Jr., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*(2), 197–206. https://doi.org/10.1086/651257

[61]. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.