福建省中电海峡智能装备研究院
Fujian Zhongdian Straits Institute of Intelligent Equipment
APSP Asia Pacific Science Press
Advances in Management and Intelligent Technologies
ISSN(O):3079-5869
ISSN(P):3079-5850
Vol. 2 No.2 (2026)

Explainability, Human Oversight, and Procedural Justice in AI-Assisted Promotion Decisions: An Integrative Review for Chinese Organizations

Peng Zhang, Rozaini Binti Rosli*
School of Business & Management, Lincoln University College, Petaling Jaya 47301, Selangor, Malaysia
Corresponding author: Rozaini Binti Rosli, rozaini@lincoln.edu.my

Abstract: AI tools are increasingly used to support internal talent decisions, yet promotion decisions pose a distinct governance problem because they involve future potential, not only past performance. Existing research has concentrated on recruitment screening or model accuracy, while the combined role of explainability, human oversight, and procedural justice in promotion contexts remains less settled. This paper develops an integrative review of research across human resource management, information systems, human-computer interaction, and AI governance to examine how managers and employees may respond to AI-assisted promotion decisions, with particular attention to Chinese organizations. Four conclusions emerge. First, explanations can improve perceived transparency, but they do not automatically protect users from poor AI advice. Second, human oversight only adds value when managers have both the authority and the criteria to question model output. Third, fairness in promotion decisions depends on voice, correctability, relevance of data, and accountability, rather than on statistical performance alone. Fourth, the Chinese regulatory context places additional emphasis on transparent and fair automated decision-making, which makes documentation, review, and appeal mechanisms especially important. On that basis, the paper proposes a practical framework for responsible AI-assisted promotion decisions built around data governance, interpretable evidence, structured human review, and employee contestability. The central argument is that organizations should not aim for uncritical trust in AI. They should aim for disciplined, reviewable, and job-relevant use.
Keywords: Explainable AI; Human Oversight; Procedural Justice; Promotion Decisions; Chinese Organizations; HR Analytics
Published: Apr 27, 2026
DOI: https://doi.org/10.62177/amit.v2i2.1341

1. Introduction

Artificial intelligence is no longer confined to recruiting chatbots or résumé screening. Across the human resource cycle, organizations are experimenting with AI to support performance management, retention analysis, capability mapping, succession planning, and internal mobility decisions[1-3]. This expansion is understandable. Internal talent decisions generate large amounts of digital trace data, and firms hope that algorithmic tools will improve consistency, speed, and predictive accuracy. Yet the move from hiring to promotion raises a different set of managerial questions.

Promotion decisions are not routine classification tasks. They allocate future opportunity, signal what the organization values, and affect how employees interpret merit, recognition, and upward mobility. A promotion case often combines structured indicators, such as past performance ratings or tenure, with harder-to-codify judgments about leadership, collaboration, learning agility, or readiness for a broader role. For this reason, promotion decisions are especially vulnerable to what Newman et al. describe as algorithmic reductionism: the tendency to compress a person into the variables that happen to be measurable[4].

Many discussions of AI in human resource management treat the central problem as one of technical quality. The usual questions are whether the model is accurate, whether biased variables have been removed, and whether the recommendations outperform intuitive judgment. Those questions matter, but they do not exhaust the issue. Managers asked to rely on AI-assisted promotion recommendations also want to know how the system reached its conclusion, whether relevant context was omitted, who remains responsible for the final decision, and whether the affected employee can challenge the outcome. These are questions about process as much as prediction.

The Chinese context gives these process questions additional practical importance. Digital transformation has accelerated the use of data-driven management tools, while the Personal Information Protection Law (PIPL) places explicit emphasis on transparency and fairness in automated decision-making[5]. At the same time, organizational justice research has long shown that employees evaluate decisions not only by their outcomes but also by the fairness of the procedures through which those outcomes are produced[6]. A promotion system can therefore generate resistance even when its statistical performance appears strong, if the process looks opaque, rigid, or unchallengeable.

Existing scholarship offers useful pieces of the puzzle, but the evidence is scattered. Research on explainable AI has examined trust and user understanding. Work on algorithmic HR has studied discrimination, applicant reactions, and managerial adoption. Human-computer interaction research has investigated contestability and oversight. What remains less settled is how these strands fit together when the decision at stake is an internal promotion rather than a recruitment screen.

This paper addresses that gap by asking a focused question: how do explainability, human oversight, and procedural justice jointly shape responsible use of AI-assisted promotion decisions in Chinese organizations? The discussion makes three moves. First, it shifts the center of attention from general AI acceptance to the more specific governance demands of promotion decisions. Second, it synthesizes mixed evidence on explainability and human review rather than treating either as a universal remedy. Third, it develops a practical framework for organizations that want to use AI in promotion processes without turning managerial judgment into ritual approval.

2. Review Design

This study adopts an integrative review design. That choice reflects the state of the field. Research relevant to AI-assisted promotion decisions is distributed across human resource management, organizational behavior, information systems, human-computer interaction, and AI governance. The literature is therefore conceptually rich but methodologically uneven, and it does not lend itself neatly to a narrow systematic review focused on one outcome variable.

The review concentrated on peer-reviewed journal articles and influential conference papers dealing with four overlapping themes: AI-assisted HR decisions, explainability and user understanding, human oversight and contestability, and procedural justice in algorithmic decision-making. Most of the reviewed work has appeared since 2019, which is when research on employment-related AI intensified, although a small number of earlier studies on trust in automation and organizational justice were retained because they remain foundational for the argument developed here[6,12,13].

Because empirical work on AI-assisted promotion decisions remains limited, the review also draws on adjacent HR contexts, especially recruiting, personnel selection, and algorithmic management. This is justified when the underlying mechanism is transferable, for example when the concern is overreliance on automated output, the effect of explanations on user judgment, or the role of voice and correctability in perceived fairness. At the same time, the paper treats promotion as a distinct case. Internal candidates carry longer organizational histories, and promotion decisions are judged not only as selection choices but also as statements about future trust and internal legitimacy.

The aim of the review is interpretive rather than bibliometric. It does not estimate pooled effect sizes or claim exhaustive coverage of every study in the field. Instead, it identifies where the evidence converges, where it remains mixed, and what those patterns imply for the design and governance of AI-assisted promotion systems in Chinese organizations.

3. Results

3.1 Explainability and Managerial Sensemaking

Explainability is often presented as the obvious response to algorithmic opacity. In broad terms, that intuition has empirical support. Studies in explainable AI show that explanations can improve perceived understanding, trust, and willingness to engage with algorithmic output[7,8]. There is, however, an important qualification. What users say they want is not always what they prioritize when stakes become concrete. Evidence also suggests that people may value interpretability in principle but still prefer accuracy when they see the two as competing objectives[9].

For promotion decisions, the function of explanation is not limited to reassurance. Managers need explanations because they must convert model output into organizationally defensible judgment. A promotion recommendation becomes usable only when managers can see which job-relevant features drove the recommendation, how strong the evidence is, what the model cannot observe, and whether the case falls inside or outside the system's intended range. In this sense, explainability supports managerial sensemaking rather than mere interface friendliness.

The literature also shows why this distinction matters. Explanations do not automatically improve decision quality. In a personnel selection task, Cecil et al. found that explainability did not offset the negative impact of incorrect AI advice[10]. Reviews of the relationship between explainability and fairness reach a similar conclusion: the fairness benefits of explainable AI are real in some settings, but they are conditional, limited, and often overstated when treated as a stand-alone solution[11]. A plausible-sounding explanation can create a veneer of rationality around a recommendation that is still based on weak data, unstable proxies, or conceptually thin measures.

This is especially problematic in promotion contexts because promotion decisions involve future potential as well as past record. If the model relies heavily on indicators that are easy to digitize, such as tenure, appraisal scores, or recent output, then even a clear explanation may still be explaining the wrong thing. In high-stakes settings, Rudin argues that organizations should prefer models that are interpretable by design rather than black-box systems explained after the fact[22]. That point is particularly relevant to promotion decisions, where a polished post hoc explanation may do less practical good than a modest model whose boundaries are intelligible.

The implication is straightforward. Useful explanations in promotion systems should be local, job-related, and bounded. They should identify the evidence that influenced the recommendation, indicate the level of uncertainty, and make explicit when human review is required. Explanations that merely increase apparent sophistication or confidence are of limited value. What organizations need is calibrated understanding, not decorative transparency.

3.2 Human Oversight beyond Symbolic Review

Organizations often respond to concerns about HR algorithms by insisting that a human manager remains in the loop. On paper, that sounds reassuring. In practice, the phrase covers very different arrangements. A manager may meaningfully interrogate the recommendation, compare it with contextual evidence, and record a reasoned final judgment. Just as easily, the manager may simply confirm the output because the system is seen as objective, fast, or politically safer than visible discretion.

The trust-in-automation literature helps explain why symbolic oversight is not enough. Appropriate reliance depends on whether users understand what the system can and cannot do, how reliable it is under different conditions, and where its failure modes lie[12,13]. Without that grounding, human review can drift toward two opposite but equally unhelpful patterns. One is algorithm aversion, where users reject the system after visible mistakes[14]. The other is passive overreliance, where the apparent neutrality of automated output discourages careful scrutiny[15].

Evidence from fairness research points in the same direction. Yurrita et al. show that oversight and contestability shape fairness perceptions when they create a genuine opportunity to question, revise, or appeal a decision[16]. Similarly, Neumann et al. find that giving human decision-makers structured autonomy in how algorithmic recommendations are used can improve perceptions and even predictive validity[17]. In other words, human involvement matters most when it is substantive enough to affect the process, not when it serves as a formal signature on a pre-determined result.

Applied to promotion decisions, meaningful oversight requires at least three conditions. First, managers must have authority to depart from the recommendation when relevant evidence justifies doing so. Second, that departure must be guided by explicit criteria, otherwise the organization simply reintroduces bias through undocumented intuition. Third, confirmations and overrides should be logged. This creates accountability in both directions: managers cannot rubber-stamp the algorithm, but they also cannot ignore it without reason.

Human oversight, then, should be treated as a governance arrangement rather than a moral slogan. The real question is not whether a person is nominally present. It is whether the person has the information, authority, and procedural obligation necessary to turn AI from a hidden decision-maker into a reviewable decision aid.

3.3 Procedural Justice in Promotion Decisions

Promotion decisions are unusually sensitive to procedural justice because they affect status, identity, pay progression, and access to future leadership roles. Organizational justice research has consistently shown that people care about voice, consistency, accuracy of information, correctability, and ethical treatment, not only about whether the final outcome benefits them personally[6]. In internal promotion contexts, these concerns are amplified because employees usually know something about the candidates, the organizational history, and the informal work that may never appear in a formal dataset.

This helps explain why fairness objections to AI-assisted HR decisions cannot be reduced to questions of statistical parity. Newman et al. argue that a system can reduce one kind of bias yet still feel unfair if it narrows the person into a thin bundle of measurable indicators[4]. The concern is not sentimental resistance to measurement. It is that the model may redefine merit in a way that excludes relational, developmental, or context-specific contributions that matter in actual organizational life.

Promotion decisions provide many examples of this risk. It is relatively easy to encode past appraisal scores, absenteeism, certification history, or sales output. It is much harder to encode whether a candidate stabilized a troubled team, mentored junior staff, absorbed extra coordination work during a transition, or demonstrated leadership under ambiguous conditions. Those elements are not always suitable for algorithmic measurement, but removing them from the decision frame can still distort what the organization means by readiness and merit.

Systematic reviews of algorithmic decision-making in HR show that discrimination concerns often stem from proxy variables, historical bias, weak job relevance, and poor data governance rather than from obviously sensitive variables alone[18]. Empirical work on algorithmic recruiting also suggests that perceived fairness depends on how transparency is communicated and how the system presents itself to users[19]. Beyond HR, studies of algorithmic management show that people interpret automated decisions through the lenses of fairness, trust, and emotion, especially when those decisions affect their status or autonomy[20].

For AI-assisted promotion decisions, procedural justice therefore requires more than a score or ranking. It requires a process in which relevant context can enter, questionable inferences can be challenged, and final responsibility remains visible. Statistical fairness metrics still matter, but they address a narrower problem. They do not tell employees whether the organization listened, reasoned carefully, or left room for correction.

3.4 Governance Challenges in Chinese Organizations

The Chinese context sharpens these issues in two ways. First, the regulatory environment gives formal weight to transparency and fairness in automated decision-making. Under Article 24 of the Personal Information Protection Law, automated decision processes are expected to be transparent and fair, especially when they materially affect individuals[5]. For organizations using AI in promotion decisions, this raises the threshold for defensibility. A model cannot be treated as acceptable merely because it performs well on internal validation metrics.

Second, AI-assisted promotion systems usually rely on cumulative employee data. These may include performance evaluations, learning records, attendance logs, competency assessments, or other forms of digital trace data. Once aggregated, such data may carry forward organizational patterns that were never intended to function as measures of promotability. Reviews of AI in HRM repeatedly note that data quality, representativeness, and governance remain unresolved problems even in organizations with strong digital ambitions[1-3].

A further complication is that promotion decisions sit at the boundary between efficiency and legitimacy. They do not merely allocate roles; they also communicate what counts as talent. Research on algorithmic management suggests that digital systems can both enable and constrain managerial autonomy[21]. The same duality is likely to appear in promotion decisions. An AI system may help standardize evaluation and flag overlooked candidates, but it may also narrow the range of acceptable reasoning if managers begin to treat the output as neutral evidence rather than as one input among several.

For Chinese organizations, responsible implementation should therefore begin with provenance and review. Managers and HR teams should know which historical decisions trained the model, which groups were underrepresented, which variables may be acting as proxies, and which cases require secondary review rather than direct adoption. In practical terms, this means that governance should focus not only on whether the model predicts something, but also on whether the recommendation can be justified under managerial scrutiny and, where necessary, challenged by the affected employee.

4. Discussion

The reviewed literature supports a simple but demanding proposition: responsible AI-assisted promotion decisions depend on the alignment of four layers of governance. The first is data relevance. The variables used in the model must have a clear job-related rationale, and the organization must be able to explain why those variables count as evidence of promotability. The second is interpretable evidence. Explanations must be specific enough to support managerial reasoning, not merely broad enough to soothe skepticism. The third is meaningful human review. Managers must be able to confirm, question, or override the recommendation according to documented criteria. The fourth is employee contestability. Individuals affected by the decision need a path to receive reasons, present omitted context, and trigger review where appropriate.

These layers reinforce one another. Explanations without contestability are one-way communication. Human oversight without criteria is informal discretion. Data governance without managerial capability produces blind dependence on technical teams. Conversely, when the four layers are aligned, AI can support disciplined promotion deliberation without displacing accountability. The goal is not to maximize trust in the system. It is to create calibrated trust: enough confidence to use the tool seriously, and enough skepticism to question it when the case demands that.

One practical implication is that organizations should resist deploying promotion models as stand-alone ranking devices. Their most defensible use is often narrower: surfacing comparable cases, flagging unusual patterns, or structuring panel discussion around explicitly defined criteria. A recommendation score may help focus attention, but promotion should still require a written rationale that links the recommendation to job demands and contextual evidence. In that sense, prediction and justification should remain analytically distinct even when they interact in practice.

A second implication concerns managerial capability. The relevant skill is not deep technical fluency in model architecture. Managers do not need to become machine-learning engineers. What they do need is the ability to ask disciplined questions. Which variables mattered most in this case? What data were missing? How stable is the recommendation? Under what conditions does the model perform poorly? What evidence justifies an override? This is a more practical form of AI literacy for promotion governance than abstract familiarity with technical jargon.

A third implication is organizational. Appeal and review mechanisms should not be treated as optional extras added after deployment. They are part of what makes the process fair in the first place. Employees are more likely to accept adverse decisions when they believe that relevant evidence can still be heard and that the process remains open to correction. Over time, this affects not only individual reactions but also the credibility of digital HR systems as a whole.

Table 1. Governance Design Priorities for AI-Assisted Promotion Decisions

Governance lever	Practical question	Minimal organizational practice	Main risk if absent
Data relevance and provenance	Why were these variables chosen as evidence of promotability?	Document job relevance, data source, update cycle, and proxy-risk checks before deployment.	Historically biased or weakly relevant signals are treated as merit indicators.
Interpretable evidence	Can managers see what drove the recommendation and where uncertainty remains?	Provide case-level explanations, uncertainty cues, and clear escalation triggers for unusual cases.	Explanations become cosmetic and create false confidence.

Governance lever	Practical question	Minimal organizational practice	Main risk if absent
Structured human review	Who can challenge or override the recommendation, and on what grounds?	Use documented review criteria, require written reasons for overrides and confirmations, and retain an audit trail.	Managers rubber-stamp the system or reintroduce bias through undocumented discretion.
Employee contestability	Can the affected employee request reasons and present missing context?	Offer explanation, review, and appeal channels for materially significant decisions.	Adverse decisions are seen as opaque, final, and procedurally unfair.
Accountability and audit	Who owns the final decision and monitors system effects over time?	Assign decision ownership, review outcome patterns, and periodically test for drift or group disparities.	Responsibility becomes diffused between HR, managers, and vendors.

4.1 Limitations and Directions for Future Research

This paper is conceptual and has two clear limits. First, the empirical literature on promotion-specific uses of AI remains thin, which means that some arguments here draw on adjacent evidence from recruiting, personnel selection, and algorithmic management. That transfer is theoretically defensible, but it is not a substitute for direct evidence from promotion settings. Second, the paper emphasizes governance in Chinese organizations without claiming that all Chinese firms face identical institutional or organizational conditions. Private firms, state-owned enterprises, and multinational subsidiaries may differ materially in how they define merit, document decisions, and manage employee appeals.

Future research should therefore move in three directions. One is to examine how middle managers actually use AI-assisted recommendations during promotion deliberations rather than asking only whether they trust the system in principle. Another is to test which forms of explanation help managers calibrate judgment without creating false confidence. A third is to compare how review and appeal rights shape employee reactions after adverse promotion outcomes. These questions are particularly important in Chinese organizations, where digital management capability and compliance expectations are both developing rapidly.

5. Conclusion

This paper has argued that the governance challenge in AI-assisted promotion decisions is not solved by predictive accuracy alone. Explainability helps, but its value depends on whether it improves understanding rather than appearance. Human oversight matters, but only when it is structured and consequential. Procedural justice remains central because employees judge promotion decisions as social and moral processes, not as optimization problems.

For Chinese organizations, these issues carry both managerial and compliance significance. Systems used in promotion decisions should be transparent enough to review, limited enough to interpret, and contestable enough to correct. AI can support more disciplined talent decisions, but only when the organization treats algorithmic output as evidence to be examined rather than as a verdict to be obeyed.

Funding

No

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] 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

[2] Budhwar, P., Malik, A., De Silva, M. T. T., & Thevisuthan, P. (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

[3] Prikshat, V., Islam, M., Patel, P., Malik, A., Budhwar, P., & Gupta, S. (2023). AI-augmented HRM: Literature review and a proposed multilevel framework for future research. Technological Forecasting and Social Change, 193, 122645. https://doi.org/10.1016/j.techfore.2023.122645

[4] 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

[5] Standing Committee of the National People's Congress. (2021). Personal Information Protection Law of the People's Republic of China, Article 24.

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

[7] 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

[8] Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The impact of explainable artificial intelligence on users' information processing. Information Systems Research, 34(4), 1582-1602. https://doi.org/10.1287/isre.2023.1199

[9] Nussberger, A.-M., Luo, L., Celis, L. E., & Crockett, M. J. (2022). Public attitudes value interpretability but prioritize accuracy in artificial intelligence. Nature Communications, 13, 5821. https://doi.org/10.1038/s41467-022-33417-3

[10] Cecil, J., Lermer, E., Hudecek, M. F. C., Sauer, J., & Gaube, S. (2024). Explainability does not mitigate the negative impact of incorrect AI advice in a personnel selection task. Scientific Reports, 14, 9736. https://doi.org/10.1038/s41598-024-60220-5

[11] Deck, L., Schoeffer, J., De-Arteaga, M., & Kühl, N. (2024). A critical survey on fairness benefits of explainable AI. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24) (pp. 1579-1595). https://doi.org/10.1145/3630106.3658990

[12] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

[13] Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407-434. https://doi.org/10.1177/0018720814547570

[14] Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114-126. https://doi.org/10.1037/xge0000033

[15] 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

[16] Yurrita, M., Draws, T., Balayn, A., Murray-Rust, D., Tintarev, N., & Bozzon, A. (2023). Disentangling fairness perceptions in algorithmic decision-making: The effects of explanations, human oversight, and contestability. In CHI 2023 - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Article 134, 1-21. https://doi.org/10.1145/3544548.3581161

[17] Neumann, M., Niessen, A. S. M., Linde, M., Tendeiro, J. N., & Meijer, R. R. (2024). 'Adding an egg' in algorithmic decision making: Improving stakeholder and user perceptions, and predictive validity by enhancing autonomy. European Journal of Work and Organizational Psychology, 33(3), 245-262. https://doi.org/10.1080/1359432X.2023.2260540

[18] Kochling, 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

[19] Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. Information Systems Journal, 34(2), 384-414. https://doi.org/10.1111/isj.12482

[20] Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1), 1-16. https://doi.org/10.1177/2053951718756684

[21] 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

[22] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x