Journal of Enterprise Information Management

Beyond efficiency: interactional foundations of fairness, accountability and transparency in AI-supported performance evaluation

Md Irfanuzzaman Khan and Robin Ladwig
Faculty of Business, Government and Law, University of Canberra, Canberra, Australia

Received 27 March 2026
Revised 22 July 2026
18 August 2026
Accepted 18 August 2026

Abstract

Purpose – This study examines how features of AI-supported performance evaluation relate to managers' judgements of fairness, accountability and transparency (FAT), and how these judgements relate to trust.

Design/methodology/approach – Survey data were collected from 289 Australian managers using a standardised performance-evaluation scenario. The hypothesised model, indirect effects, moderation effects, alternative specifications and out-of-sample prediction were assessed using PLS-SEM.

Findings – Interaction quality, human agency and perceived humanness were positively associated with all three FAT dimensions, whereas voice was associated only with transparency. All three FAT dimensions were positively associated with trust, with transparency showing the largest coefficient. Most specific indirect effects were significant, but task complexity and perceived uncanniness did not moderate the FAT–trust relationships. Human agency showed slightly greater unique explanatory value than voice, while the overall FAT appraisal in the exploratory collective model was strongly associated with trust.

Research limitations/implications – The cross-sectional research design identifies associations rather than causal effects, while the Australian, AI-experienced sample limits generalisability.

Practical implications – Organisations should combine understandable interaction with contextual sensitivity, human review, documented override authority, traceable responsibility and credible appeal mechanisms.

Originality/value – This study brings together human–AI interaction factors to explain trust in AI-supported evaluations. It shows that fairness, accountability and transparency make distinct contributions to trust, while their combined contribution can also be represented as an overall FAT appraisal.

Keywords Artificial intelligence, Performance evaluation, Organisational justice, Fairness, Accountability, Transparency, Human agency, Trust

Paper type Research article

1. Introduction

Artificial intelligence (AI) has expanded enterprise information systems beyond the automation of simple and repetitive tasks. Supported by advances in big data, algorithms and computational power, AI-based systems can sense, analyse and respond to complex organisational information in ways that augment traditional human decision support (Zdravković and Panetto, 2022). However, the value organisations derive from these applications depends on how effectively technological capabilities are aligned with organisational arrangements and the demands of the task being performed (Agarwal, 2023; Tinguely *et al.*, 2023). As AI becomes embedded within everyday enterprise processes, it also redistributes decision authority across managers, data infrastructures, software interfaces and external vendors. Consequently, considerations of fairness, control, responsibility, and contestability are central to enterprise information management (Lippert *et al.*, 2026; Tarafdar *et al.*, 2023; Wang *et al.*, 2025).

Journal of Enterprise Information Management
Emerald Publishing Limited
e-ISSN: 1758-7409
p-ISSN: 1741-0398
DOI 10.1108/JEIM-03-2026-0476

This redistribution of authority is consequential in performance management, as AI progressively informs how employees are assessed, developed and managed (Pan *et al.*, 2026). AI-supported systems draw on multiple forms of employee and performance data to generate scores, explanations and recommendations. These outputs can enhance consistency and expand the information available to managers, although performance appraisal still requires contextual judgement and consideration of future employee development (Varma *et al.*, 2024). Employees are also sensitive to whether feedback comes from a person or an algorithm and to the kinds of information used to generate an evaluation (Biswas *et al.*, 2024; Majrashi, 2025; Qin *et al.*, 2023). The managerial challenge is therefore to determine whether an AI-generated recommendation provides a sufficiently credible basis for making a consequential judgement about another person.

Extant literature offers constructive insights into this challenge through studies of accuracy, explainability, bias, and differences between human and machine decision-makers (Köchling and Wehner, 2020; Langer and Landers, 2021). However, the evidence is mixed. In some contexts, AI appears more consistent or fair than human judgement, yet opaque or reductive procedures can raise significant justice concerns (Choung *et al.*, 2024; Newman *et al.*, 2020; Wesche *et al.*, 2024). Studies in employee selection, gig work, and algorithmic management indicate that individuals evaluate the overall decision-making process as well as the final outcome. Their assessments are influenced by the information used, the explanations provided, and the degree of human involvement (Acikgoz *et al.*, 2020; Jabagi *et al.*, 2025; Lee, 2018; Yu *et al.*, 2025). However, there is limited understanding of how these factors interact when managers employ AI to evaluate employees.

Research on algorithm aversion and appreciation provides further insight into this issue. Individuals may reject algorithmic advice after observing an error, become more willing to use an imperfect algorithm when permitted to modify its recommendations, or prefer algorithmic judgement in suitable decision contexts (Dietvorst *et al.*, 2015, 2018; Logg *et al.*, 2019). Recent evidence also positions trust as central to how HR professionals evaluate AI tools. A study of 324 Swiss HR professionals found that trust increased the perceived usefulness of AI-enabled HR tools, with perceived decision fairness partially mediating this relationship (Revillod, 2025). Holistically, these studies indicate that managers' trust in AI-supported performance evaluation goes beyond the technical quality of an AI-generated recommendation. It is also likely to depend on whether the process appears fair, responsibility for the decision remains clear and the basis of the recommendation is understandable. We accordingly ask:

RQ. How are the interactional features of AI-supported performance evaluation associated with managers' judgements of fairness, accountability and transparency, and how do these judgements relate to trust?

To address this question, we bring organisational justice theory into conversation with research on fairness, accountability and transparency (FAT). Organisational justice explains how individuals judge the legitimacy of outcomes, procedures, information and interpersonal treatment, while FAT dimensions extend these concerns into algorithmic settings by focusing on impartiality, understandability and visible responsibility (Colquitt, 2001; Greenberg, 1987; Shin, 2020; Shin and Park, 2019). We do not treat FAT as a substitute for established justice dimensions. Instead, FAT dimensions are viewed as justice-relevant appraisals through which managers interpret an AI-supported decision process. We examine four observable features as potential antecedents of these appraisals: interaction quality, human agency, perceived humanness and voice. From a holistic standpoint, these features constitute a coherent set of interactional and decision-governance cues, as they signal whether the system communicates clearly, acknowledges relevant human context, allows for input, and remains subject to meaningful human judgement (Cui *et al.*, 2024; Fanni *et al.*, 2023; Hellwig *et al.*, 2023; Koh *et al.*, 2025; Schuetzler *et al.*, 2020).

The empirical analysis is conducted in two stages. The initial conceptual model examines how the four interactional features relate to fairness, accountability and transparency, and how these appraisals subsequently relate to trust. Task complexity and perceived uncanniness are also investigated as potential boundary conditions. The alternative models then compare the unique contributions of voice and human agency and assess whether the combined contribution of fairness, accountability and transparency can be represented as an overall FAT appraisal. Accordingly, this study makes three focused contributions. First, it extends the application of organisational justice reasoning to AI-supported performance evaluation by showing that interaction quality, perceived humanness and retained human agency are consistently associated with managers' fairness, accountability and transparency judgements, which in turn are associated with trust. Second, it refines the concept of human involvement by distinguishing voice from agency. Voice is associated only with transparency, whereas agency is associated with all three FAT dimensions and retains slightly greater unique explanatory value in reduced-model comparisons. Third, the study clarifies how FAT can be examined at two analytical levels. The initial model treats fairness, accountability and transparency as distinct appraisals that contribute separately to trust, while the exploratory alternative model represents their combined contribution as a FAT appraisal.

2. Theoretical background

2.1 AI-supported performance evaluation

AI-supported performance evaluation integrates employee data and analytical models into managerial appraisal and decision-making processes. These systems generate performance scores, explanations, and recommendations; however, their organisational consequences depend on system design, integration with existing practices, and utilisation by those responsible for employment decisions (Bankins, 2021; Bujold *et al.*, 2024; Tinguely *et al.*, 2023). Research on AI-supported selection and performance feedback indicates that both managers and employees respond to the manner in which AI is incorporated into decision processes. Decision-makers may assign disproportionate importance to AI-generated rankings, while employees differentiate between AI, human, and combined feedback, and evaluate the perceived fairness of the assessor (Biswas *et al.*, 2024; Malin *et al.*, 2024; Qin *et al.*, 2023). These factors are significant when evaluations influence employee development, recognition, and career advancement. We therefore distinguish between an AI-generated evaluation and an AI-supported performance evaluation. The former refers to the score, explanation, or recommendation produced by the system, while the latter encompasses the organisational process in which managers review, interpret, contextualise, and act upon the output. The present study focuses on AI-supported performance evaluation, as the legitimacy of the final judgement depends on how the system is integrated into managerial decision-making.

2.2 Organisational justice and FAT

Organisational justice theory offers a framework for understanding how managers assess the legitimacy of AI-supported performance evaluations. Distributive justice pertains to the appropriateness of outcomes, while procedural justice addresses the consistency, accuracy, correctability, and ethicality of the evaluation process. Informational justice relates to the adequacy and truthfulness of explanations, and interpersonal justice concerns the dignity and respectful treatment of individuals (Bies and Moag, 1986; Colquitt, 2001; Leventhal *et al.*, 1980; Thibaut and Walker, 1975). These dimensions suggest that confidence in an evaluation depends on how it is generated, explained, and communicated, in addition to the outcome itself. However, applying these principles is challenging when the underlying algorithmic process is only partially visible. Opacity may result from technical complexity or intentional secrecy, and organisational disclosures often provide limited information about the collection

and use of personal data (Burrell, 2016; Crain, 2018). Consequently, managers may find it difficult to understand how employee information is selected, weighted, and translated into recommendations. Such evaluations may reduce complex performance and contextual factors to a single score or prediction (Binns *et al.*, 2018; Newman *et al.*, 2020). Managers are therefore required to interpret the process based on observable features of the system and decision-making arrangement.

The proposed framework outlined in Figure 1 emphasises four key features of the human–AI decision arrangement. Interaction quality assesses whether the system communicates clearly, responds appropriately and effectively supports the evaluation task. Voice refers to the extent to which managers can provide relevant information before an evaluation is finalised, aligning with evidence that opportunities for voice influence responses to automated decisions (Hellwig *et al.*, 2023). Human agency denotes the retained authority to challenge, correct or override a recommendation, based on research linking agency to contestability and redress (Fanni *et al.*, 2023). Perceived humanness concerns whether the interaction appears socially recognisable and human-like, including apparent recognition of employee needs and feedback resembling that of a human manager (Lu *et al.*, 2022). Although the four features are conceptually distinct, they are unified by a common principle, as each provides observable information about whether a partially opaque evaluation process can be understood, contextualised, influenced, and corrected. Consequently, we argue that these features constitute a coherent group of interactional and decision-governance signals, rather than an exhaustive list of all possible antecedents of FAT.

The above observable features do not demonstrate that the underlying algorithm is unbiased or ensure a just outcome. Instead, organisational justice provides a basis for evaluating aspects of the decision process, including opportunities for input and correction, access to adequate information and fair treatment (Colquitt, 2001; Fanni *et al.*, 2023), whereas FAT captures managers' assessments of whether AI-supported decisions are fair, accountable

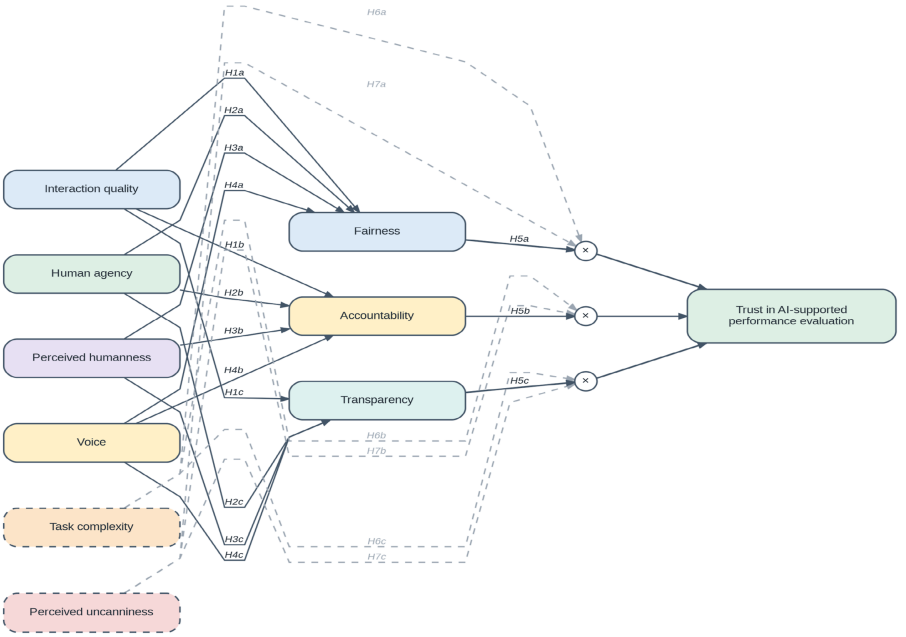

Figure 1. Conceptual model. Source: Authors' own work

and transparent (Shin and Park, 2019). We therefore treat interaction quality, perceived humanness, voice and human agency as justice-relevant process cues that managers may draw on when forming these FAT judgements. Consequently, organisational justice explains why these process features are relevant, while FAT captures managers' AI-specific appraisals of the resulting decision process.

Thus, managers' FAT judgements concern three distinct aspects of an AI-supported appraisal process. Fairness addresses whether the evaluation appears impartial and appropriate; transparency pertains to whether the information and reasoning underlying the recommendation are comprehensible; and accountability concerns whether responsibility and justification remain identifiable (Diakopoulos, 2016; Shin, 2020; Shin and Park, 2019). We do not treat the three FAT dimensions as direct substitutes for the established organisational justice dimensions. Instead, we view fairness, accountability and transparency as AI-specific appraisals of the decision process that are informed by the justice-relevant cues described above. This conceptualisation aligns with related work on responsible AI. Explainable-AI research examines methods for making opaque models more intelligible, while algorithmic-accountability research emphasises answerability, traceability and identifiable responsibility (Adadi and Berrada, 2018; Diakopoulos, 2016).

Recent trustworthy artificial intelligence (AI) frameworks prioritise fairness, transparency and accountability alongside broader considerations such as human oversight, privacy, inclusivity and system robustness (Hosseini Tabaghdehi and Ayaz, 2025; Kowald *et al.*, 2024). The initial conceptual model analyses FAT dimensions as separate constructs, recognising that each addresses a distinct concern. For example, an evaluation may be perceived as fair while the basis for the recommendation remains unclear, or transparent while responsibility for the final decision remains ambiguous (Kowald *et al.*, 2024; Shin and Park, 2019). The proposed alternative model assesses whether the three dimensions can also be represented collectively as an overall FAT appraisal. This specification complements the original model and functions as a robustness check rather than introducing a new construct or replacing the three individual appraisals.

2.3 Trust in AI-supported performance evaluation

Trust is especially significant in AI-supported performance evaluation because managers must assess recommendations that they may be unable to verify fully yet remain accountable for the outcomes. Lee and See (2004) define trust as an attitude that an automated agent will assist in achieving the user's goals under conditions of uncertainty and vulnerability. In this context, trust refers to a manager's confidence that an AI-supported recommendation offers a credible foundation for evaluating another employee. It is important to distinguish managers' trust in the system from their reliance on a specific recommendation. Trust denotes evaluative confidence in the system, whereas reliance concerns the behavioural decision to accept or implement its recommendation (Hoff and Bashir, 2015; Klingbeil *et al.*, 2024). For example, a manager may trust the system's overall capabilities but reject an evaluation that fails to consider relevant contextual information. Appropriate trust should therefore be calibrated to the system's demonstrated strengths and limitations while preserving managerial scrutiny and responsibility (Lee and See, 2004; Lucas *et al.*, 2024).

FAT dimensions offer complementary foundations for establishing this confidence. Fairness enables managers to evaluate whether the assessment is impartial and appropriate. Transparency allows managers to understand the rationale behind the recommendation and to identify any limitations in the information or reasoning applied. Accountability ensures that responsibility, justification, and opportunities for review are clearly defined. Previous research links FAT appraisals to trust in algorithmic systems (Shin, 2020; Shin and Park, 2019). Collectively, these factors assist managers in determining whether an AI-supported recommendation constitutes a sufficiently credible basis for performance evaluation. Figure 1 outlines the study's initial overarching conceptual model.

3. Hypothesis development

3.1 Interaction quality and FAT perceptions

Interaction quality encompasses clarity, responsiveness, usability, and task support experienced by users when engaging with an AI system. Research on chatbot design demonstrates that conversational structure, disclosure, responsiveness, and socially appropriate interaction influence users' interpretation of exchanges, especially as task complexity increases (Chaves and Gerosa, 2021; Cheng *et al.*, 2022; Haugeland *et al.*, 2022). Within organisational communication, conversational design can indicate listening and transparency (Men *et al.*, 2022). Additionally, anthropomorphic and fairness cues collectively influence user satisfaction with AI agents (Koh *et al.*, 2025).

From an organisational-justice perspective, these features extend beyond enhancing usability. Clear information enables managers to determine whether appropriate criteria have been applied. Responsive interaction demonstrates that the system can receive and process decision-relevant input, rather than merely providing a predetermined output. Research on humanised chatbot design and explainability further indicates that interaction design affects both user experience and the perceived understandability and acceptability of system decisions (Rhim *et al.*, 2022; Shin, 2021).

Interaction quality may also influence perceived accountability. When queries, feedback and the sequence of the evaluation are visible, managers can better understand how the recommendation was handled and where responsibility for the final decision remains. Thus, meaningful human oversight requires people to retain a supervisory and decision-making role and to be able to explain and justify how AI is used within the wider process (Kowald *et al.*, 2024). Interaction quality may therefore make managerial involvement and answerability more visible, leading us to expect a positive association with accountability.

H1a-c. Interaction quality is positively associated with perceptions of (a) fairness, (b) accountability and (c) transparency in AI-supported performance evaluation.

3.2 Human agency and FAT perceptions

Human agency refers to the practical capacity to interpret, question, contextualise and override an AI-generated recommendation. It is more than the visible presence of a human reviewer: agency requires meaningful decision rights and the ability to exercise them. Contestability and redress give this capacity operational form by allowing a person to challenge data, reasoning or outcomes (Fanni *et al.*, 2023). Experimental evidence also suggests that a stronger sense of agency is associated with greater confidence and acceptability in human–machine interaction (Vantrepotte *et al.*, 2022). More broadly, research on AI fairness and over-reliance highlights the risks of placing too much weight on automated decisions (Narayanan *et al.*, 2024; Spatola, 2024). Recent research on human oversight also emphasises that the reliable detection of inaccurate or unfair AI outputs is essential for ensuring effective oversight (Langer *et al.*, 2024).

This reasoning follows the procedural-justice emphasis on correctability and control. A decision process appears more legitimate when inaccurate or incomplete information can be challenged before the outcome is finalised (Leventhal *et al.*, 1980; Thibaut and Walker, 1975). In the context of performance appraisal, recorded data may exclude temporary personal circumstances, collaborative contributions, or qualitative aspects of work that managers consider relevant (Bankins *et al.*, 2022; Newman *et al.*, 2020). Retaining human agency allows managers to bring this contextual knowledge into the evaluation instead of relying exclusively on the AI-generated assessment. When managers can review and override AI outputs, responsibility is less likely to become obscured within the system. Human oversight facilitates the translation of technical information into organisationally intelligible reasons, enables the identification of limitations, and links the final decision to an accountable role (Bovens, 2010; Kowald *et al.*, 2024). We therefore expect human agency to strengthen FAT perceptions.

H2a-c. Human agency is positively associated with perceptions of (a) fairness, (b) accountability and (c) transparency in AI-supported performance evaluation.

3.3 Perceived humanness and FAT perceptions

Building on previous research on perceived humanness (Lu *et al.*, 2022; Shin, 2022), this study defines perceived humanness as the degree to which interactions with AI are experienced as human-like. In the context of performance evaluation, the construct focuses on managers' perceptions of human-like communication and relational cues, without suggesting that the system possesses authentic human emotions or qualities (Porra *et al.*, 2020). This operationalisation aligns with Lu *et al.* (2022), who conceptualise perceived humanness through conversational human voice, anthropomorphism, and social presence. Prior research also indicates that human-like conversational and relational cues can enhance perceptions of social presence and anthropomorphism, and can influence trust, emotional assurance, and acceptance of AI-based interactions (Pelau *et al.*, 2021; Schuetzler *et al.*, 2020; Shin, 2022).

In the performance-appraisal context, managers are required to deliver challenging feedback, consider exceptional circumstances and maintain the dignity of the employee under evaluation. An AI system perceived as recognising employee needs and communicating in a human-like manner may indicate that the appraisal process extends beyond merely assigning a numerical score. Anthropomorphic communication can make AI interactions appear more human-like and may facilitate acceptance (Pelau *et al.*, 2021). In AI-supported performance evaluation, such relational cues may also inform fairness judgements by influencing how appropriately the process is perceived to operate (Bankins *et al.*, 2022; Bies and Moag, 1986). Perceived humanness may further support transparency and accountability by making AI-generated reasoning easier for managers to interpret and communicate. Accordingly, we expect perceived humanness to be positively associated with all three FAT dimensions.

H3a-c. Perceived humanness is positively associated with perceptions of (a) fairness, (b) accountability and (c) transparency in AI-supported performance evaluation.

3.4 Voice and FAT perceptions

Voice refers to the opportunity to provide relevant information or express concerns before a decision is finalised. It is a well-established element of procedural justice because people value being heard even when they do not control the outcome (Colquitt, 2001; Lind *et al.*, 1990; Thibaut and Walker, 1975). In AI-supported evaluation, however, managers may still be unsure whether their input has been understood or meaningfully considered. Hellwig *et al.* (2023) show that providing an opportunity for voice can improve perceived fairness. More recent evidence indicates that voice can also increase perceived control over automated decision processes, while remaining conceptually distinct from final decision authority (Hellwig *et al.*, 2025). Voice must therefore be distinguished from human agency. The distinction reflects Thibaut and Walker's (1975) separation between process control and decision control. Voice allows managers to contribute information, whereas agency gives them the authority to question, correct or override the recommendation. A manager may therefore be heard without having any real influence over the outcome (Fanni *et al.*, 2023).

Drawing on this reasoning, we expect voice to strengthen perceptions of FAT in AI-supported performance evaluation. We also argue that a visible exchange of information can make the evaluation easier to follow and clarify the manager's involvement, thereby supporting perceptions of transparency and accountability. These benefits, however, are more likely to emerge when the opportunity to contribute is genuine rather than merely symbolic (Fanni *et al.*, 2023; Hellwig *et al.*, 2023). Thus, we propose the following hypotheses:

H4a-c. Voice is positively associated with perceptions of (a) fairness, (b) accountability and (c) transparency in AI-supported performance evaluation.

3.5 FAT perceptions and trust

Research on algorithm aversion and algorithm appreciation shows that people are neither consistently resistant nor automatically receptive to algorithmic advice. They may reject an algorithm after seeing it make an error, even when it remains more accurate than human judgement (Dietvorst *et al.*, 2015). However, resistance decreases when users retain some ability to modify the algorithm's recommendation (Dietvorst *et al.*, 2018). Other studies indicate that people may prefer algorithmic advice in appropriate judgement tasks (Logg *et al.*, 2019). These contrasting findings indicate that accuracy is only one part of the judgement managers make when deciding whether to trust an AI recommendation. Lee (2018), for example, found that algorithmic managerial decisions were trusted less when the task required intuition and human judgement. Qin *et al.* (2023), however, found that AI could be perceived as fairer and more accurate in structured, data-intensive evaluations. Related studies show that explanations and the distribution of authority between humans and AI also shape how algorithmic decisions are assessed (Köchling *et al.*, 2025; Shulner-Tal *et al.*, 2025).

Christin (2017) further shows that algorithms gain influence through how organisational actors interpret and use them in practice. Thus, trust is influenced not only by algorithmic recommendations but also by the interaction between managers and artificial intelligence systems. As discussed earlier, fairness, accountability, and transparency each serve as distinct foundations for trust in algorithmic systems. Consequently, when the process appears fair, understandable and answerable, managers are more likely to trust the evaluation while still exercising critical judgement (Shin, 2020; Shin and Park, 2019). Thus, we propose the following hypotheses:

H5a-c. Perceptions of (a) fairness, (b) accountability and (c) transparency are positively associated with trust in AI-supported performance evaluation.

3.6 Moderating role of task complexity

Task complexity reflects the number of interdependent details, concentration demands and difficult-to-track stages involved in a decision (Wood, 1986). A complex performance evaluation may require managers to integrate quantitative results, qualitative evidence and changing contextual information. Research in technology-mediated service settings suggests that complexity changes how users evaluate disclosure, problem-solving capability and automated support (Cheng *et al.*, 2022; Xu *et al.*, 2020), while performance-management research highlights the broader complexity of organisational measurement systems (Okwir *et al.*, 2018). We therefore model task complexity as a possible boundary condition on the relationship between each FAT dimension and trust. Competing model analysis later considers a different possibility: complexity may shape the formation of FAT appraisals rather than alter their subsequent relationship with trust. Thus, we propose the following hypotheses:

H6a-c. Task complexity negatively moderates the relationships between (a) fairness, (b) accountability and (c) transparency and trust in AI-supported performance evaluation.

3.7 Moderating role of perceived uncanniness

The uncanny valley literature suggests that greater human likeness does not always produce more favourable reactions. When human-like features are combined with noticeably artificial or inconsistent behaviour, the resulting mismatch can evoke feelings of eeriness or discomfort (Ho and MacDorman, 2010; Mori *et al.*, 2012). Our study does not test the full nonlinear uncanny valley relationship. Instead, perceived uncanniness captures whether the AI system feels strange, unsettling or inappropriately human-like within a consequential performance-evaluation setting. Such feelings may act as a negative affective cue when managers decide whether to trust the evaluation. Prior research shows that feelings of uncanniness can weaken

trust in AI systems (Sullivan *et al.*, 2022). We therefore argue that, even when an evaluation appears fair, accountable and transparent, these positive appraisals may translate less strongly into trust when the system itself feels unsettling. Thus, we propose the following hypotheses:

H7a-c. Perceived uncanniness negatively moderates the relationships between (a) fairness, (b) accountability and (c) transparency and trust in AI-supported performance evaluation.

4. Method

4.1 Participants and procedure

Data were collected from 289 managers across Australia through an online research panel specialising in managerial and organisational samples. Participants were screened using four self-reported inclusion criteria intended to identify respondents with relevant experience of AI-enabled managerial processes. They were asked whether they knew enough about AI to engage with it effectively, whether they had used AI in their work, whether they had used or were currently using AI for performance management, and how often they used AI technology, with response options ranging from daily to never. Only respondents who met the self-reported familiarity and use criteria were invited to complete the full survey. Participants then completed the measures after reading a standardised AI-supported performance evaluation scenario. Table 1 presents the demographic profile of the respondents.

The final sample consisted of 289 Australian managers. Most were Millennials (54.7%) or Generation Z (23.2%), followed by Generation X (19.0%) and Baby Boomers (2.8%). The majority worked in the private sector (68.5%), with additional representation from the public (22.1%) and not-for-profit sectors (4.5%). Medium-sized organisations were most common (54.7%), alongside large firms (30.4%) and small businesses (14.9%). In terms of hierarchy, 36.3% were middle managers, 23.9% supervisors or frontline leaders, 12.5% senior leaders

Table 1. Participant demographic characteristics ($N = 289$)

Variable	Category	Frequency (*n*)	Percentage (%)
Generation	Generation Z (1997–2012)	67	23.2
	Millennials (1981–1996)	158	54.7
	Generation X (1965–1980)	55	19.0
	Baby Boomers (1946–1964)	8	2.8
	Silent Generation (1925–1945)	1	0.3
Sector	Private	198	68.5
	Public	64	22.1
	Not-for-profit	13	4.5
	Other/Missing	14	4.8
Organisation size	Small (<20 employees)	43	14.9
	Medium (20–199 employees)	158	54.7
	Large (200+ employees)	88	30.4
Level	Middle Management	105	36.3
	Supervisor/Frontline Leader	69	23.9
	Senior Leader	36	12.5
	Executive	24	8.3
	Other (Specialist, 2iC, etc.)	55	19.0
Team size	Small (<6 members)	53	18.3
	Medium (6–15 members)	160	55.4
	Large (16+ members)	76	26.3
Working arrangement	Hybrid	187	64.7
	On-site	65	22.5
	Remote	37	12.8

and 8.3% executives, with the remainder in specialist, second-in-command or other roles. Most operated in hybrid work arrangements (64.7%), followed by on-site (22.5%) and fully remote settings (12.8%).

4.2 Measures

All constructs were measured using five-point Likert scales ranging from 1 (strongly disagree) to 5 (strongly agree). The measurement items were adapted from established measures where appropriate and otherwise contextualised for the AI-supported performance-evaluation context using prior conceptual and empirical work. Interaction quality items were contextualised based on the research on human–AI interaction quality and interaction design (Haugeland *et al.*, 2022; Pelau *et al.*, 2021). Human agency draws on work on sense of agency, contestability and redress (Fanni *et al.*, 2023; Vantrepotte *et al.*, 2022). Perceived humanness was adapted from Lu *et al.*'s (2022) validated perceived-humanness measure and contextualised to performance appraisal. Voice was adapted from Colquitt's (2001) process-control and voice items and contextualised to automated decision-making using Hellwig *et al.* (2023). Fairness was based on organisational-justice measurement and research on perceived algorithmic fairness (Colquitt, 2001; Shin, 2020; Shin and Park, 2019), while accountability was informed by work on responsibility and answerability in organisational and algorithmic settings (Bovens, 2010; Shin and Park, 2019). Transparency draws on work on understandability, explainability and observability in algorithmic decision-making (Shin, 2020, 2021; Shin and Park, 2019). Trust was grounded in research on trust in automation and algorithmic systems (Lee and See, 2004; Shin and Park, 2019). Task complexity followed Wood's (1986) conceptualisation, while perceived uncanniness was informed by established research on eeriness and uncanniness in artificial agents (Ho and MacDorman, 2010; Sullivan *et al.*, 2022). Before estimating the conceptual model, FAIR3 (I trust the AI system to make fair decisions) was removed. Although originally intended to capture fairness, its explicit reference to trust risked blurring the distinction between fairness as an antecedent and trust as the outcome. Fairness was therefore estimated using FAIR1 and FAIR2, consistent with its conceptualisation as perceived impartiality and freedom from bias (Colquitt, 2001; Shin, 2020). Table 2 presents the measurement items, standardised loadings, reliability and convergent-validity statistics, together with the measurement basis for each construct.

4.3 Analytical strategy

We selected PLS-SEM to estimate a relatively complex variance-explanation model with multiple antecedents, parallel mediating mechanisms, interaction terms and a formative collective-FAT alternative (Becker *et al.*, 2012; Hair *et al.*, 2019). The choice reflected our interest in explained variance, indirect effects and the stability of conclusions across theoretically plausible specifications. We used SEMinR 2.5.0, the path-weighting scheme and reflective Mode A measurement for the first-order constructs. Statistical inference relied on 5,000 nonparametric case-resampling bootstrap draws and two-tailed percentile confidence intervals. We assessed the measurement model through indicator loadings, Cronbach's alpha, composite reliability, average variance extracted (AVE), the Fornell–Larcker criterion and the heterotrait–monotrait ratio of correlations (HTMT) (Fornell and Larcker, 1981; Henseler *et al.*, 2015). Structural assessment included inner variance inflation factors (VIF), path coefficients, confidence intervals, R^2, adjusted R^2 and f^2. We addressed common-method risk procedurally through anonymity, neutral instructions, scenario standardisation and separate construct blocks. As supplementary diagnostics, the first unrotated component accounted for 40.0% of retained-indicator variance and full-collinearity VIF values ranged from 1.14 to 3.11. These results provide no indication that a single general method factor dominates the covariance structure.

We also compared theory-consistent alternatives. Such comparisons are widely used as robustness checks in structural modelling because they reveal whether a conclusion depends

Table 2. Measurement items and measurement-model statistics

Construct	Item	Measurement items	Loading	Alpha	CR	AVE	Measurement basis
Interaction quality	IQ1	The AI system provides clear information during performance evaluations	0.795	0.827	0.885	0.658	Pelau *et al.* (2021), Haugeland *et al.* (2022)
	IQ2	The AI is easy to use for HR tasks	0.836				
	IQ3	The AI system responds quickly to feedback	0.795				
	IQ4	AI makes HR processes easier to manage	0.818				
Human agency	HA1	I feel in control of the AI-assisted evaluation process	0.784	0.688	0.826	0.613	Fanni *et al.* (2023), Vantrepotte *et al.* (2022)
	HA2	I can provide feedback to the AI system	0.818				
	HA3	I can override the AI decisions if necessary	0.745				
Perceived humanness	PHUM1	The AI understands employee needs during performance reviews	0.852	0.830	0.898	0.746	Lu *et al.* (2022)
	PHUM2	The AI system provides feedback like a human manager	0.867				
	PHUM3	Interaction with the AI feels like dealing with a person	0.872				
Voice	VOICE1	I can express my opinions during AI-led performance evaluations	0.836	0.728	0.846	0.647	Colquitt (2001), Hellwig *et al.* (2023)
	VOICE2	My input is considered in the AI system's decisions	0.819				
	VOICE3	I feel I can influence AI decisions that affect me	0.755				
Fairness	FAIR1	The AI system treats employees fairly in HR processes	0.890	0.700	0.869	0.769	Colquitt (2001), Shin and Park (2019), Shin (2020)
	FAIR2	I believe the AI system is unbiased	0.864				
Accountability	ACC1	The AI system holds me accountable for my performance	0.805	0.717	0.841	0.638	Bovens (2010), Shin and Park (2019)
	ACC2	I am responsible for the work in the AI system	0.749				
	ACC3	The AI decisions are explainable and justifiable	0.840				

(*continued*)

Table 2. Continued

Construct	Item	Measurement items	Loading	Alpha	CR	AVE	Measurement basis
Transparency	TRANS1	I understand how the AI system makes HR decisions	0.815	0.792	0.878	0.707	Shin and Park (2019), Shin (2020, 2021)
	TRANS2	The AI system explains its HR decision-making process	0.854				
	TRANS3	The AI provides enough information about its decisions	0.853				
Trust in AI	TRUST1	I trust the AI to evaluate employees' performance accurately	0.854	0.787	0.876	0.702	Lee and See (2004), Shin and Park (2019)
	TRUST2	The AI system follows through on its promises	0.829				
	TRUST3	Overall the decision by AI can be trusted	0.829				
Task complexity	TC1	The HR task using AI involves managing many details	0.820	0.686	0.829	0.619	Wood (1986)
	TC2	Using AI for HR decisions requires concentration	0.822				
	TC3	AI tasks in HR involve several steps that are difficult to track	0.712				
Perceived uncanniness	UNC1	The AI system feels unnatural when used for HR decisions	0.875	0.849	0.905	0.762	Ho and MacDorman (2010), Sullivan *et al.* (2022)
	UNC2	The AI's human-like qualities make me uncomfortable	0.909				
	UNC3	The AI feels unsettling in HR process	0.832				

Note(s): FAIR3 was removed before final estimation because it explicitly incorporated trust wording. All reported analyses use FAIR1 and FAIR2

on one arrangement of the constructs (Danks *et al.*, 2020; Sarstedt *et al.*, 2020). Model 1 is the original differentiated structure. Model 2A removes voice but retains human agency; Model 2B removes human agency but retains voice. Comparing the two reveals how much unique explanatory information is lost when one form of human participation is excluded. We also assessed indicator-level out-of-sample prediction using PLSpredict with ten folds and ten repetitions, comparing PLS prediction errors against a linear-model benchmark.

Model 3 is the collective FAT model, described methodologically as a higher-order specification. Its substantive question is whether fairness, accountability and transparency, while judged separately, can also be represented collectively as a broader appraisal of the AI-supported evaluation arrangement. Following Becker *et al.* (2012), we estimated the model with a disjoint two-stage approach using the existing first-order construct scores rather than new survey items. Model 3 retained interaction quality, human agency and perceived humanness and, as an exploratory extension, included task complexity and uncanniness as

antecedents of the collective FAT judgement. The alternative models are robustness analyses; they do not replace the original hypotheses.

5. Results

5.1 Preliminary diagnostics and measurement model

We first assessed the 30 retained indicators across 289 cases. Loadings ranged from 0.712 to 0.909, composite reliability from 0.826 to 0.905 and AVE from 0.613 to 0.769. Cronbach's alpha ranged from 0.686 to 0.849. Human agency ($\alpha = 0.688$) and task complexity ($\alpha = 0.686$) were marginally below 0.70, but both exceeded 0.82 in composite reliability and 0.61 in AVE. The two-item fairness scale produced $\alpha = 0.700$, composite reliability of 0.869 and AVE of 0.769. Taken together, the loadings, composite reliability and AVE support convergent validity, while the marginal alpha values for two short scales warrant measured interpretation.

We assessed discriminant validity using the Fornell–Larcker criterion. As shown in Table 3, the square root of AVE for each construct, reported on the diagonal, exceeded its correlations with all other constructs. The square roots of AVE ranged from 0.783 to 0.877, while the largest inter-construct correlation was 0.738 between human agency and voice. These results indicate that each construct shared more variance with its own indicators than with the other constructs in the model, providing support for discriminant validity under the Fornell–Larcker criterion (Fornell and Larcker, 1981). This pattern is also consistent with our theoretical distinction between the constructs.

Inner VIF values ranged from 1.49 to 2.76 for the four antecedent equations and from 1.88 to 2.33 for the trust equation, indicating no severe structural collinearity. Full-collinearity VIF values ranged from 1.14 to 3.11, and the first unrotated component accounted for 40.0% of total retained-indicator variance. These diagnostics indicate only that the observed covariance pattern is not dominated by a single factor. The same-source, cross-sectional design therefore remains an acknowledged limitation.

5.2 Hypothesised structural model

As outlined in Table 4 and Figure 2, interaction quality was positively associated with fairness ($\beta = 0.235$), accountability ($\beta = 0.204$) and transparency ($\beta = 0.221$). Human agency showed the same pattern for fairness ($\beta = 0.206$), accountability ($\beta = 0.250$) and transparency ($\beta = 0.254$). Perceived humanness produced the largest antecedent coefficients for fairness ($\beta = 0.286$), accountability ($\beta = 0.273$) and transparency ($\beta = 0.314$). H1a-c, H2a-c and H3a-c are therefore supported.

Voice produced a more differentiated pattern. It did not explain unique variance in fairness ($\beta = 0.130$, $p = 0.162$) or accountability ($\beta = 0.158$, $p = 0.063$), so H4a and H4b are not supported. Voice was positively associated with transparency ($\beta = 0.156$, 95% CI [0.017, 0.288], $p = 0.027$), supporting H4c. The result indicates that opportunities for input were related to whether managers understood how the evaluation was reached. They were not independently associated with perceptions that the process was unbiased or that responsibility for the decision remained clear.

FAT explained 61.1% of trust in the main-effects model (adjusted $R^2 = 0.607$). Fairness ($\beta = 0.321$), accountability ($\beta = 0.228$) and transparency ($\beta = 0.339$) were each positively associated with trust, supporting H5a–c. The effect sizes were generally small. Perceived humanness had a medium effect on transparency ($f^2 = 0.170$) and small effects on fairness ($f^2 = 0.110$) and accountability ($f^2 = 0.107$). Interaction quality and human agency produced small effects across the three FAT dimensions, ranging from 0.031 to 0.060. Voice had a negligible effect on fairness ($f^2 = 0.014$) and small effects on accountability ($f^2 = 0.022$) and transparency ($f^2 = 0.025$). Among the predictors of trust, fairness produced $f^2 = 0.140$, transparency $f^2 = 0.127$ and accountability $f^2 = 0.057$. Adding task complexity, uncanniness

Table 3. Fornell–Larcker assessment

Construct	1	2	3	4	5	6	7	8	9	10
Interaction quality	[0.811]									
Human agency	0.689	[0.783]								
Perceived humanness	0.495	0.473	[0.864]							
Voice	0.604	0.738	0.532	[0.804]						
Fairness	0.597	0.599	0.569	0.576	[0.877]					
Accountability	0.607	0.636	0.577	0.611	0.623	[0.799]				
Transparency	0.646	0.670	0.627	0.644	0.642	0.712	[0.841]			
Trust in AI	0.687	0.692	0.573	0.676	0.681	0.669	0.707	[0.838]		
Task complexity	0.649	0.557	0.478	0.567	0.505	0.616	0.543	0.534	[0.787]	
Perceived uncanniness	0.079	0.138	0.224	0.204	0.054	0.075	0.093	0.108	0.218	[0.873]

Note(s): Diagonal values in brackets are square roots of AVE. Construct order: 1 = interaction quality; 2 = human agency; 3 = perceived humanness; 4 = voice; 5 = fairness; 6 = accountability; 7 = transparency; 8 = trust; 9 = task complexity; 10 = perceived uncanniness

Table 4. Hypothesis tests for the differentiated model

Hyp	Path	Beta	SE	95% CI	*p*	Decision
H1a	Interaction quality → Fairness	0.235	0.078	[0.089, 0.394]	0.002	Supported
H1b	Interaction quality → Accountability	0.204	0.071	[0.076, 0.355]	<0.001	Supported
H1c	Interaction quality → Transparency	0.221	0.069	[0.093, 0.365]	<0.001	Supported
H2a	Human agency → Fairness	0.206	0.085	[0.044, 0.373]	0.013	Supported
H2b	Human agency → Accountability	0.250	0.091	[0.066, 0.427]	0.008	Supported
H2c	Human agency → Transparency	0.254	0.082	[0.096, 0.414]	0.001	Supported
H3a	Perceived humanness → Fairness	0.286	0.074	[0.140, 0.429]	<0.001	Supported
H3b	Perceived humanness → Accountability	0.273	0.057	[0.162, 0.384]	<0.001	Supported
H3c	Perceived humanness → Transparency	0.314	0.053	[0.212, 0.419]	<0.001	Supported
H4a	Voice → Fairness	0.130	0.090	[−0.050, 0.301]	0.162	Not supported
H4b	Voice → Accountability	0.158	0.083	[−0.008, 0.315]	0.063	Not supported
H4c	Voice → Transparency	0.156	0.069	[0.017, 0.288]	0.027	Supported
H5a	Fairness → Trust	0.321	0.057	[0.212, 0.436]	<0.001	Supported
H5b	Accountability → Trust	0.228	0.064	[0.103, 0.354]	<0.001	Supported
H5c	Transparency → Trust	0.339	0.064	[0.213, 0.460]	<0.001	Supported
H6a	Fairness × Task complexity → Trust	0.035	0.065	[−0.112, 0.151]	0.569	Not supported
H6b	Accountability × Task complexity → Trust	−0.031	0.072	[−0.162, 0.119]	0.694	Not supported
H6c	Transparency × Task complexity → Trust	0.009	0.065	[−0.123, 0.134]	0.914	Not supported
H7a	Fairness × Uncanniness → Trust	−0.113	0.070	[−0.228, 0.050]	0.166	Not supported
H7b	Accountability × Uncanniness → Trust	0.024	0.066	[-0.122, 0.139]	0.825	Not supported
H7c	Transparency × Uncanniness → Trust	−0.020	0.067	[−0.146, 0.120]	0.817	Not supported

and the six interaction terms increased explained variance in trust to 63.0% (adjusted $R^2 = 0.616$), but the additional direct moderator and interaction effect sizes were negligible, ranging from 0.000 to 0.014. Complete effect-size results are reported in Table 5.

5.3 Mediation and moderation

We evaluated specific indirect effects directly with 5,000 bootstrap resamples (see Table 6). Interaction quality, human agency and perceived humanness each had significant indirect relationships with trust through fairness, accountability and transparency. The largest effects were perceived humanness through transparency (0.107), perceived humanness through fairness (0.092) and human agency through transparency (0.086). Voice had no significant indirect relationship through fairness or accountability, but its indirect relationship through transparency was significant (effect = 0.053, 95% CI [0.005, 0.110], $p = 0.027$). Ten of the twelve specific indirect effects were therefore significant. Task complexity did not moderate any FAT–trust relationship, so H6a–c is not supported. None of the perceived-uncanniness interactions reached the two-tailed 5% threshold, providing no support for H7a–c. The evidence therefore does not support either moderator as a reliable boundary condition on the translation of FAT judgements into trust.

5.4 Competing-model assessment

The competing models were not estimated to replace Model 1 or to search for more favourable significance levels. Rather, they ask whether the same substantive conclusions survive other

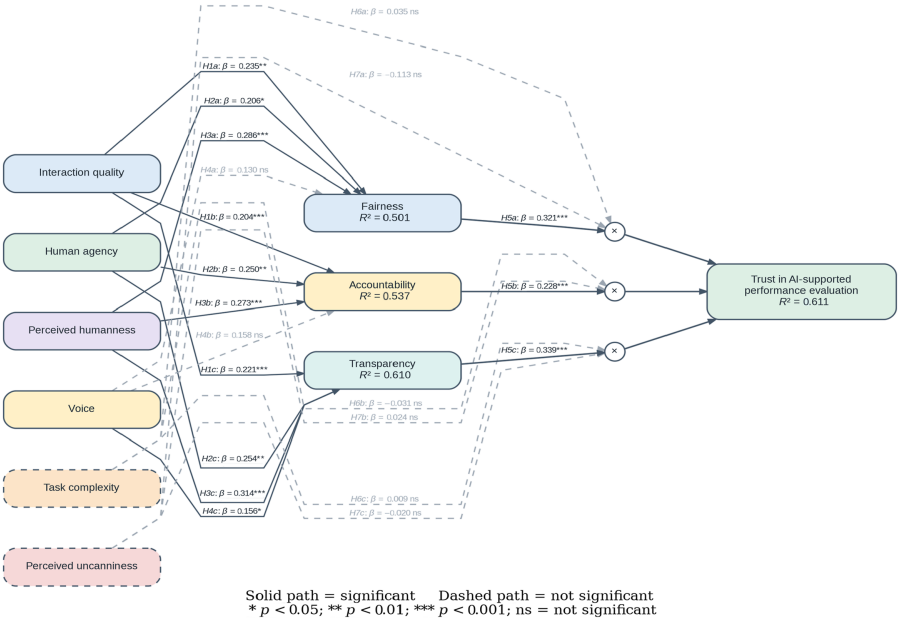

Figure 2. Estimated model. Note: Solid paths represent significant relationships and dashed paths represent non-significant relationships. R^2 values shown for the endogenous constructs correspond to the main-effects model. The dashed H6 and H7 paths report moderation estimates from the extended moderated model, in which the explained variance in trust increased to $R^2 = 0.630$. Source: Authors' own work

reasonable arrangements of the constructs, a recognised robustness practice in PLS-SEM and multimodel inference (Danks *et al.*, 2020; Sarstedt *et al.*, 2020). Model 2A removes voice but retains agency; Model 2B does the reverse. Model 3 asks whether fairness, accountability and transparency can also be treated as contributors to one overall FAT judgement.

Model 2A, which retained human agency but removed voice, explained 49.5% of fairness, 52.7% of accountability and 60.0% of transparency, only slightly below Model 1. Model 2B, which retained voice but removed human agency, explained 48.6, 51.5 and 58.7%, respectively. Trust remained at approximately 61.1% in all three differentiated models because the same three FAT dimensions predicted trust. As shown in Table 7, removing voice therefore produced small losses of 0.6 to 1.0% points across the FAT outcomes, whereas removing agency produced larger losses of 1.5 to 2.3% points. The comparison indicates that human agency offers slightly greater unique explanatory value than voice. However, the difference is modest, and voice remains relevant, particularly in promoting transparency.

Model 3 retained the Model 2A interactional antecedents and, in an explicitly exploratory extension, added task complexity and uncanniness as antecedents of the overall FAT judgement. Interaction quality ($\beta = 0.189$, $p = 0.005$), human agency ($\beta = 0.348$, $p < 0.001$), perceived humanness ($\beta = 0.361$, $p < 0.001$) and task complexity ($\beta = 0.158$, $p = 0.007$) were positively associated with overall FAT. Uncanniness was negative but did not reach the two-tailed 5% threshold ($\beta = -0.093$, $p = 0.059$). Overall FAT was strongly associated with trust ($\beta = 0.781$, $p < 0.001$). Fairness, accountability and transparency all contributed significantly to the overall FAT appraisal, with weights of 0.362, 0.329 and 0.445, respectively. The model explained 71.9% of overall FAT and 61.0% of trust.

In substantive terms, Model 3 asks whether three related judgements can be considered together without assuming that they are identical. Fairness, accountability and transparency remain separate dimensions, while their shared contribution is summarised as a broader

Table 5. Complete f^2 effect sizes for the structural models

Hypothesis	Structural path	f^2	Interpretation
Panel A. Main model			
H1a	IQ → FAIR	0.052	Small
H1b	IQ → ACC	0.043	Small
H1c	IQ → TRANS	0.060	Small
H2a	HA → FAIR	0.031	Small
H2b	HA → ACC	0.048	Small
H2c	HA → TRANS	0.060	Small
H3a	PHUM → FAIR	0.110	Small
H3b	PHUM → ACC	0.107	Small
H3c	PHUM → TRANS	0.170	Medium
H4a	VOICE → FAIR	0.014	Negligible
H4b	VOICE → ACC	0.022	Small
H4c	VOICE → TRANS	0.025	Small
H5a	FAIR → TRUST	0.140	Small
H5b	ACC → TRUST	0.057	Small
H5c	TRANS → TRUST	0.127	Small
Panel B. Additional direct moderator and interaction effects in the moderated model			
–	TC → TRUST	0.000	Negligible
–	UNC → TRUST	0.000	Negligible
H6a	FAIR × TC → TRUST	0.002	Negligible
H6b	ACC × TC → TRUST	0.001	Negligible
H6c	TRANS × TC → TRUST	0.000	Negligible
H7a	FAIR × UNC → TRUST	0.014	Negligible
H7b	ACC × UNC → TRUST	0.001	Negligible
H7c	TRANS × UNC → TRUST	0.001	Negligible

Note(s): Panel A reports every structural path in the differentiated main model. Panel B reports the additional direct moderator and interaction effects introduced in the moderated model; the main structural paths are not duplicated. Effect sizes are interpreted as negligible (<0.02), small (0.02–0.149), medium (0.15–0.349) and large (≥0.35). Values shown as 0.000 are rounded to three decimal places

Table 6. Bootstrapped specific indirect effects

Indirect path	Effect	SE	95% CI	*p*	Decision
Interaction quality → Fairness → Trust	0.075	0.029	[0.026, 0.140]	0.002	Significant
Interaction quality → Accountability → Trust	0.047	0.022	[0.013, 0.099]	<0.001	Significant
Interaction quality → Transparency → Trust	0.075	0.027	[0.029, 0.135]	<0.001	Significant
Human agency → Fairness → Trust	0.066	0.031	[0.013, 0.133]	0.013	Significant
Human agency → Accountability → Trust	0.057	0.028	[0.011, 0.120]	0.008	Significant
Human agency → Transparency → Trust	0.086	0.034	[0.027, 0.161]	0.001	Significant
Perceived humanness → Fairness → Trust	0.092	0.027	[0.043, 0.147]	<0.001	Significant
Perceived humanness → Accountability → Trust	0.062	0.020	[0.026, 0.103]	<0.001	Significant
Perceived humanness → Transparency → Trust	0.107	0.024	[0.062, 0.157]	<0.001	Significant
Voice → Fairness → Trust	0.042	0.031	[−0.015, 0.110]	0.162	Not significant
Voice → Accountability → Trust	0.036	0.023	[−0.002, 0.086]	0.063	Not significant
Voice → Transparency → Trust	0.053	0.027	[0.005, 0.110]	0.027	Significant

Table 7. Competing-model explanatory comparison

Model	R^2 fairness	R^2 accountability	R^2 transparency	R^2 overall FAT	R^2 trust	Adj. R^2 trust
M1	0.501	0.537	0.610	–	0.611	0.607
M2A	0.495	0.527	0.600	–	0.611	0.607
M2B	0.486	0.515	0.587	–	0.611	0.607
M3	–	–	–	0.719	0.610	0.609

appraisal (see Figure 3). We use this reflective-formative two-stage specification (Becker *et al.*, 2012) only to assess the robustness of the original differentiated model. The results indicate that the three dimensions can be represented collectively as an overall FAT appraisal, but the exploratory antecedent paths require independent confirmation.

5.5 Predictive and sensitivity assessment

As a supplementary predictive check, we conducted indicator-level PLSpredict using ten folds and ten repetitions and compared PLS prediction errors with a linear-model benchmark. For Model 1, average out-of-sample RMSE was 0.735 for PLS and 0.752 for the linear model, while average MAE was 0.551 and 0.560, respectively. PLS produced lower RMSE for nine of the eleven predicted indicators. Model 2A produced average RMSE of 0.737 compared with 0.751 for the benchmark, and Model 2B produced 0.742 compared with 0.758. Thus, all three models showed useful predictive performance, and Model 1 had the lowest PLS RMSE, although the differences among the models were modest.

Predictive superiority was not uniform across every outcome. Model 1 clearly improved prediction for fairness, accountability and transparency, but the trust comparison was mixed: average PLS RMSE for trust was 0.690 compared with 0.692 for the linear model, while average PLS MAE was slightly higher (0.518 versus 0.513). We therefore describe the model as having generally useful, rather than uniformly superior, out-of-sample predictive performance. Table 8 reports the indicator-level out-of-sample prediction results.

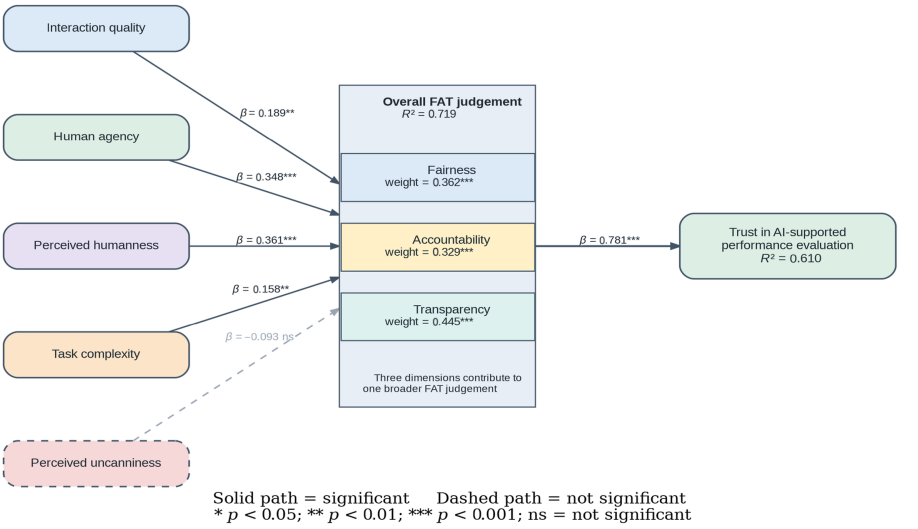

Figure 3. Alternative collective FAT model. Source: Authors' own work

Table 8. Indicator-level out-of-sample prediction

Model	Outcome	PLS RMSE	LM RMSE	ΔRMSE	PLS MAE	LM MAE	Lower PLS RMSE
M1	FAIR	0.760	0.772	−0.012	0.570	0.578	2/2
M1	ACC	0.768	0.784	−0.016	0.578	0.589	3/3
M1	TRANS	0.730	0.766	−0.036	0.542	0.567	3/3
M1	TRUST	0.690	0.692	−0.002	0.518	0.513	1/3
M1	Overall	0.735	0.752	−0.017	0.551	0.560	9/11
M2A	FAIR	0.761	0.769	−0.008	0.573	0.579	2/2
M2A	ACC	0.772	0.784	−0.012	0.581	0.589	3/3
M2A	TRANS	0.733	0.759	−0.026	0.548	0.564	3/3
M2A	TRUST	0.691	0.697	−0.006	0.519	0.516	2/3
M2A	Overall	0.737	0.751	−0.013	0.554	0.561	10/11
M2B	FAIR	0.764	0.769	−0.005	0.574	0.577	1/2
M2B	ACC	0.774	0.794	−0.020	0.582	0.598	3/3
M2B	TRANS	0.742	0.771	−0.028	0.548	0.570	3/3
M2B	TRUST	0.693	0.701	−0.008	0.520	0.518	2/3
M2B	Overall	0.742	0.758	−0.016	0.554	0.565	9/11

Note(s): Ten-fold PLSpredict with ten repetitions. Values are averages across the indicators belonging to each outcome. LM = linear-model benchmark; ΔRMSE = PLS RMSE minus LM RMSE, so negative values favour PLS. The final column reports the number of indicators for which PLS produced a lower RMSE than the benchmark

6. Discussion

6.1 Interactional antecedents of FAT judgements

The findings show a differentiated pattern across the four antecedents. Interaction quality, human agency and perceived humanness were positively associated with fairness, accountability and transparency and were indirectly associated with trust through all three dimensions. Voice followed a narrower path, as it was associated only with transparency and indirectly with trust through transparency.

Interaction quality reflects the overall experience of working with an AI-supported evaluation system, including its clarity, responsiveness, ease of use and task support. When decision processes are not fully visible, these observable qualities may help managers judge whether the process is understandable and appropriately conducted. This interpretation is consistent with fairness heuristic theory, which suggests that people rely on procedural cues when forming justice and trust judgements under uncertainty (Lind, 2001). It also complements research showing that interaction structure and responsiveness influence evaluations of digital systems (Chaves and Gerosa, 2021; Men *et al.*, 2022). We therefore interpret interaction quality as the quality of managers' engagement with the AI-supported evaluation system. This experience was associated with all three FAT dimensions and is consistent with algorithmic-management research that treats outcomes as emerging from human–algorithm interactions and their organisational context (Lippert *et al.*, 2026; Tarafdar *et al.*, 2023).

The difference between human agency and voice leads to a more detailed conclusion. Voice is important in procedural-justice theory since it enables people to state their opinions and take part in the decision-making process (Colquitt, 2001; Thibaut and Walker, 1975). In our study, though, voice was linked only to transparency. This indicates that giving people an opportunity to contribute and having some influence may help managers to see how an evaluation was arrived at and whether their input was taken into account, without necessarily affecting people's perception of AI-supported decisions. This pattern means that the usual effects of voice in relation to AI-assisted decisions should be qualified because simply being heard may mainly improve the visibility of information when input is not coupled with the authority to

change the recommendation. The justice value of voice may therefore be stronger when opportunities for process control are accompanied by meaningful influence over the eventual decision (Colquitt, 2001; Thibaut and Walker, 1975). This interpretation is supported by recent evidence which shows that voice can increase perception of control over automated decision processes (Hellwig *et al.*, 2025).

On the other hand, human agency, which captured control and the ability to override AI decisions, was associated with all three FAT dimensions. This pattern is consistent with research emphasising contestability, redress and meaningful human involvement in algorithmic decisions (Fanni *et al.*, 2023; Langer *et al.*, 2024; Spatola, 2024). The reduced-model comparisons point in the same direction, although the differences were modest. Removing human agency produced slightly larger reductions in the explained variance of fairness, accountability and transparency than removing voice, while explained variance in trust remained almost unchanged. We therefore do not treat agency and voice as substitutes. Voice concerns the opportunity to contribute information and have it considered; agency additionally concerns the authority to question, correct or override the recommendation.

Perceived humanness showed the largest antecedent coefficients across the three FAT dimensions. The measure captured whether the AI appeared to recognise employee needs, provided feedback resembling that of a human manager and felt human-like during the interaction. We do not consider these perceptions as evidence that AI systems possess empathy, nor do we treat them as a justification for simply imitating human behaviour. Instead, we argue that managers may evaluate AI-supported processes more favourably when the interaction is socially recognisable and appears responsive to relevant employee needs, while human judgement remains essential for interpreting and acting on AI-generated outputs. This interpretation is also consistent with interactional justice, which emphasises respectful and appropriate treatment (Bies and Moag, 1986), and with research showing that human-like qualities can improve evaluations of AI when they make interactions more understandable and socially meaningful (Koh *et al.*, 2025; Lu *et al.*, 2022; Pelau *et al.*, 2021). It is also consistent with Shin *et al.* (2025), who found that AI involvement in HR decisions can contribute to organisational dehumanisation and negative employee reactions. Our study did not measure dehumanisation, but the findings indicate that perceived human-like communication remains relevant to managers' FAT judgements.

6.2 Task complexity and perceived uncanniness

Neither task complexity nor perceived uncanniness moderate the relationships between fairness, accountability and transparency and trust. All six interaction effects were non-significant and negligible in size. The original conceptual model therefore provides little support for the proposed boundary conditions. We do not interpret these findings as suggesting that contextual factors are unimportant. Instead, they indicate greater complexity in determining where such boundary conditions may operate. We argue that some contextual influences may enter earlier in the judgement process, affecting how managers form their FAT appraisals rather than altering the subsequent relationship between those appraisals and trust. Once managers have formed judgements about FAT dimensions, their relationships with trust may be comparatively stable across the levels of task complexity and uncanniness observed in this study. Task complexity, for example, may affect how closely managers examine available information, explanations and opportunities for human intervention while forming these judgements. This interpretation is consistent with extant research which highlights the relevance of task- and system-level conditions for detecting problematic AI outputs (Langer *et al.*, 2024). The collective model also provides tentative support for this earlier-stage interpretation, as task complexity was positively associated with overall FAT. Because this relationship was not hypothesised *a priori*, however, we treat it as exploratory.

Uncanniness may similarly function as an initial affective response rather than as a moderator. Sullivan *et al.* (2022) position uncanniness as a response that can precede reduced trust. In our collective model, however, its negative relationship with overall FAT remained non-significant. We therefore found no clear evidence for either its moderating or earlier direct role. Future experimental research could manipulate task complexity and human-like design to determine where these conditions enter the formation of FAT and trust judgements.

6.3 Differentiated and collective FAT judgements

Fairness, accountability and transparency were each positively associated with trust. Transparency had the largest coefficient, followed closely by fairness and then accountability, although these differences should not be treated as a definitive ranking. Fairness captured whether the AI-supported process was considered fair and unbiased; transparency concerned understanding and explanation; and accountability reflected responsibility and the justifiability of decisions. Accountability should therefore be interpreted within the evaluation process rather than as broader legal or institutional accountability.

The transparency finding is consistent with research linking understandable and explainable AI processes with trust (Shin, 2020; Shin and Park, 2019). However, transparency should not be treated as a substitute for fairness or accountability. Mirbabaie *et al.* (2026) similarly found that transparency did not uniformly improve all forms of perceived fairness. Our differentiated model therefore supports treating the three dimensions as related but substantively distinct grounds for trust.

The collective competing model outlined in Figure 3 provides a complementary robustness test. Fairness, accountability and transparency each contributed to the overall FAT appraisal, which was significantly associated with trust and explained almost the same trust variance as the original model. This does not show that managers consciously combine the three dimensions or that the collective model is superior. It shows that their combined contribution can be represented as an overall FAT appraisal without materially changing the central trust result. We therefore maintain both analytical levels. The original model provides greater diagnostic detail, while the collective model recognises the combined contribution of FAT dimensions.

7. Theoretical implications

We contribute to research into AI-supported performance evaluation by showing how organisational justice and FAT dimensions provide distinct but complementary perspectives on managers' evaluations of AI-supported decisions. Organisational justice explains why features of procedures, information and interpersonal treatment are important. On the other hand, FAT captures managers' AI-specific appraisals of fairness, accountability and transparency. Integrating these perspectives clarifies how observable interactional and decision-governance cues inform FAT judgements, which in turn relate to trust in AI-supported performance evaluation. Prior studies have examined justice perceptions in algorithmic HR decisions (Acikgoz *et al.*, 2020; Bankins *et al.*, 2022; Qin *et al.*, 2023), while related research has considered how human-like system qualities influence responses to AI (Lu *et al.*, 2022; Pelau *et al.*, 2021). We do not claim to extend organisational justice theory itself. Rather, we extend its application to AI-enabled HR information systems by specifying how interactional and decision-governance cues are associated with FAT judgements.

The findings show that interaction quality, human agency and perceived humanness were each positively associated with fairness, accountability and transparency. Perceived humanness showed the largest coefficients across the three dimensions, although the differences do not establish a definitive ranking of effects. The social and interactional qualities of an HR information system therefore become part of the information managers use when forming FAT judgements (Koh *et al.*, 2025; Pelau *et al.*, 2021).

The contrast between voice and human agency adds further nuance. Voice is important in procedural justice because it allows affected individuals to provide information and participate in a decision process (Colquitt, 2001; Hellwig *et al.*, 2023, 2025; Thibaut and Walker, 1975). In our study, however, voice was associated only with transparency, whereas human agency was associated with all three FAT dimensions. Removing human agency produced a somewhat greater, but still modest, loss of explanatory power than removing voice. The findings therefore suggest cautiously that allowing users to provide input may help them understand how an evaluation was reached and whether their input was considered, without necessarily assuring them that the recommendation can be challenged, corrected or changed. Retained human authority appears to carry justice implications beyond the availability of an input channel (Dietvorst *et al.*, 2018; Fanni *et al.*, 2023).

We also show that FAT can be examined at both differentiated and collective levels. The differentiated model captures the distinct contribution of each dimension, while the collective model represents their combined contribution as an overall FAT appraisal. These specifications therefore provide different analytical views of FAT in AI-supported performance evaluation.

8. Practical implications

Because interaction quality, human agency and perceived humanness were each associated with FAT dimensions, AI-supported performance appraisal should be managed as an enterprise decision infrastructure rather than merely as an analytical tool. Organisations should designate a clearly identified accountable owner for each system, document its intended purpose and permitted uses, specify the responsibilities of human resources, line managers, information technology teams and external vendors, and establish explicit criteria for deployment and periodic review. An organisation-wide AI register should record the system's data sources, capabilities, limitations, risk assessments, test results and accountable decision-makers. These practices align with contemporary guidance on accountable ownership, AI registers, lifecycle monitoring and human control, as well as the management-system approach established by ISO/IEC 42001 (International Organization for Standardization, 2023; National AI Centre, 2026).

The relatively strong relationships involving perceived humanness highlight contextual recognition as an important design priority. Interfaces should communicate in a humane manner, specify the information used to generate recommendations and acknowledge missing, uncertain or potentially outdated evidence. Managers should be able to add relevant contextual information before finalising an evaluation. The system should clearly disclose that recommendations are AI-generated and avoid presenting probabilistic outputs as definitive managerial conclusions. This approach supports human-centred interaction and reduces the risk that human-like communication may blur the distinction between AI-generated outputs and human judgement (Hosseini Tabaghdehi and Ayaz, 2025; National AI Centre, 2026).

The distinction between voice and human agency offers more specific guidance for human oversight. Voice was associated exclusively with transparency, whereas human agency was linked to all three dimensions of FAT and showed slightly greater explanatory value in the comparative models. This suggests that opportunities for input may be more meaningful when managers also retain sufficient authority to question, correct or escalate AI-generated recommendations. Thus, organisations should ensure that managers possess adequate information, training, time, and authority to question recommendations, request additional evidence, correct errors, pause evaluations, and override or escalate outputs. Opportunities for input should be traceable; the interface should indicate whether contextual evidence influenced the recommendation and explain why it was retained if no change occurred. Australian guidance underscores meaningful human control, contestability, and accessible redress, while the EU AI Act serves as an international benchmark by requiring competent overseers who can interpret, disregard, override, or reverse high-risk AI outputs (European Parliament and Council of the European Union, 2024; National AI Centre, 2026).

Given that FAT dimensions were each independently associated with trust, organisations should monitor them as distinct operational qualities. Fairness reviews should assess whether errors or adverse outcomes are disproportionately distributed across employee groups and whether relevant contextual information is consistently considered. Transparency assessments should determine whether managers understand the information used, the model's limitations, and the appropriate role of its recommendations. Accountability requires designated decision owners, documented reviews and overrides, accessible appeal procedures, and clear records of corrective actions. Maintaining decision logs, audit records, and regular review cycles enables these qualities to be observable and assessable, rather than relying on general assertions of system responsibility (Cheong, 2024; International Organization for Standardization, 2023; National AI Centre, 2026).

From the perspective of enterprise information management, AI-supported performance appraisal constitutes a component of a broader organisational decision-making process. An AI system may generate a performance score or recommendation, after which managers review the underlying information, consider relevant employee circumstances, interpret the output, and determine whether to accept, question, or override it. The quality of the resulting decision depends on information management practices, the integration of human judgement, and the allocation of responsibility for the final decision. This perspective aligns with research indicating that algorithmic systems function within broader human and organisational contexts (Bujold *et al.*, 2024; Lippert *et al.*, 2026; Tarafdar *et al.*, 2023), as well as studies highlighting the complementary role of human judgement in conjunction with algorithmic capabilities (Cui *et al.*, 2024). Our findings suggest that managers' FAT judgements are associated with the quality of interaction with the AI system, perceived humanness, and the extent of human agency retained in the evaluation process.

These recommendations should be regarded as implementation priorities rather than established causal interventions. The cross-sectional, same-source design identifies features associated with managers' evaluations of AI-supported appraisal but does not demonstrate that introducing specific practices will necessarily increase trust. Organisations should therefore pilot these practices, collect feedback from managers and affected employees, and evaluate whether they improve decision quality, understanding, and appropriate human oversight within their own context.

9. Limitations and future research

Several limitations qualify our conclusions. First, the cross-sectional design identifies theory-consistent associations but cannot establish causality. Longitudinal studies should examine how trust changes after managers observe errors, override recommendations or receive evidence about employee outcomes. Experiments could also manipulate explanation quality, voice and override authority to clarify their distinct effects. Second, the survey ensured consistent appraisal context but could not reproduce the complexity of a live AI-enabled HR information system. Field research should combine perceptual measures with system logs, actual overrides, appeal outcomes, model confidence and employee reactions. Third, the sample was drawn from an online panel and relied on self-reported eligibility. Participants' actual use of AI in performance management was not independently verified and representativeness of Australian managers cannot be established. The sample was also relatively young and highly engaged with AI, which further limits generalisability. Cross-national and cross-sector studies should examine whether regulation, labour institutions and prior discrimination experience shape FAT judgements (Moritz and Wehner, 2026). Fourth, the accountability measure captures a bounded, decision-level form of answerability within the immediate human–AI evaluation arrangement. It reflects whether responsibility remains identifiable and whether an AI-supported decision can be explained and justified, but it does not capture the wider legal, regulatory or institutional architecture of algorithmic accountability. Future research

should therefore use multi-level measures that distinguish managerial responsibility, organisational accountability, vendor responsibility and formal regulatory oversight. Fifth, future studies should use measures that distinguish distributive, procedural, informational and interpersonal dimensions without incorporating trust language. Experimental research should also manipulate task complexity and human-like system design to determine whether they influence the formation of FAT appraisals, their relationship with trust, or both. Finally, the collective FAT model was a theory-informed exploratory alternative rather than a preregistered confirmatory model and therefore requires independent replication. Future research should also test whether differentiated FAT judgements and an overall FAT appraisal predict different outcomes, including reliance, contestation, employee acceptance and managerial responsibility.

10. Conclusion

We began this study by asking how specific features of AI-supported performance evaluation are associated with managers' fairness, accountability and transparency judgements, and how these judgements relate to trust. Interaction quality, human agency and perceived humanness were each positively associated with all three FAT dimensions, whereas voice was associated only with transparency. Fairness, accountability and transparency were subsequently associated with trust, with transparency producing the largest coefficient. Task complexity and perceived uncanniness did not moderate the FAT–trust relationships. Within the limitations of a cross-sectional design, our findings suggest that trust in AI-supported performance evaluation is associated with clear interaction, human-centred communication, meaningful human authority and managers' judgements that the process is fair, accountable and transparent.

Acknowledgments

The authors would like to thank Tim Millar for his assistance and support during the development of this research.

References

Acikgoz, Y., Davison, K.H., Compagnone, M. and Laske, M. (2020), "Justice perceptions of artificial intelligence in selection", *International Journal of Selection and Assessment*, Vol. 28 No. 4, pp. 399-416, doi: 10.1111/ijsa.12306.

Adadi, A. and Berrada, M. (2018), "Peeking inside the black-box: a survey on explainable artificial intelligence", *IEEE Access*, Vol. 6, pp. 52138-52160, doi: 10.1109/access.2018.2870052.

Agarwal, A. (2023), "AI adoption by human resource management: a study of its antecedents and impact on HR system effectiveness", *Foresight*, Vol. 25 No. 1, pp. 67-81, doi: 10.1108/fs-10-2021-0199.

Bankins, S. (2021), "The ethical use of artificial intelligence in human resource management: a decision-making framework", *Ethics and Information Technology*, Vol. 23 No. 4, pp. 841-854, doi: 10.1007/s10676-021-09619-6.

Bankins, S., Formosa, P., Griep, Y. and 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*, Vol. 24 No. 3, pp. 857-875, doi: 10.1007/s10796-021-10223-8.

Becker, J.-M., Klein, K. and Wetzels, M. (2012), "Hierarchical latent variable models in PLS-SEM: guidelines for using reflective-formative type models", *Long Range Planning*, Vol. 45 Nos 5-6, pp. 359-394, doi: 10.1016/j.lrp.2012.10.001.

Bies, R.J. and Moag, J.S. (1986), "Interactional justice: communication criteria of fairness", in Lewicki, R.J., Sheppard, B.H. and Bazerman, M.H. (Eds), *Research on Negotiation in Organizations*, JAI Press, Greenwich, CT, Vol. 1, pp. 43-55.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J. and Shadbolt, N. (2018), "'It's reducing a human being to a percentage': perceptions of justice in algorithmic decisions", *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Paper 377, pp. 1-14.

Biswas, M.I., Talukder, M.S. and Khan, A.R. (2024), "Who do you choose? Employees' perceptions of artificial intelligence versus humans in performance feedback", *China Accounting and Finance Review*, Vol. 26 No. 4, pp. 512-532, doi: 10.1108/cafr-08-2023-0095.

Bovens, M. (2010), "Two concepts of accountability: accountability as a virtue and as a mechanism", *West European Politics*, Vol. 33 No. 5, pp. 946-967, doi: 10.1080/01402382.2010.486119.

Bujold, A., Roberge-Maltais, I., Parent-Rocheleau, X., Boasen, J., Sénécal, S. and Léger, P.-M. (2024), "Responsible artificial intelligence in human resources management: a review of the empirical literature", *AI and Ethics*, Vol. 4 No. 4, pp. 1185-1200, doi: 10.1007/s43681-023-00325-1.

Burrell, J. (2016), "How the machine 'thinks': understanding opacity in machine learning algorithms", *Big Data and Society*, Vol. 3 No. 1, doi: 10.1177/2053951715622512.

Chaves, A.P. and Gerosa, M.A. (2021), "How should my chatbot interact? A survey on social characteristics in human-chatbot interaction design", *International Journal of Human-Computer Interaction*, Vol. 37 No. 8, pp. 729-758, doi: 10.1080/10447318.2020.1841438.

Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022), "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure", *Internet Research*, Vol. 32 No. 2, pp. 496-517, doi: 10.1108/intr-08-2020-0460.

Cheong, B.C. (2024), "Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making", *Frontiers in Human Dynamics*, Vol. 6, 1421273, doi: 10.3389/fhumd.2024.1421273.

Choung, H., Seberger, J.S. and David, P. (2024), "When AI is perceived to be fairer than a human: understanding perceptions of algorithmic decisions in a job application context", *International Journal of Human-Computer Interaction*, Vol. 40 No. 22, pp. 7451-7468, doi: 10.1080/10447318.2023.2266244.

Christin, A. (2017), "Algorithms in practice: comparing web journalism and criminal justice", *Big Data and Society*, Vol. 4 No. 2, doi: 10.1177/2053951717718855.

Colquitt, J.A. (2001), "On the dimensionality of organizational justice: a construct validation of a measure", *Journal of Applied Psychology*, Vol. 86 No. 3, pp. 386-400, doi: 10.1037/0021-9010.86.3.386.

Crain, M. (2018), "The limits of transparency: data brokers and commodification", *New Media and Society*, Vol. 20 No. 1, pp. 88-104, doi: 10.1177/1461444816657096.

Cui, T., Tan, B. and Shi, Y. (2024), "Fostering humanistic algorithmic management: a process of enacting human-algorithm complementarity", *The Journal of Strategic Information Systems*, Vol. 33 No. 2, 101838, doi: 10.1016/j.jsis.2024.101838.

Danks, N.P., Sharma, P.N. and Sarstedt, M. (2020), "Model selection uncertainty and multimodel inference in partial least squares structural equation modeling (PLS-SEM)", *Journal of Business Research*, Vol. 113, pp. 13-24, doi: 10.1016/j.jbusres.2020.03.019.

Diakopoulos, N. (2016), "Accountability in algorithmic decision making", *Communications of the ACM*, Vol. 59 No. 2, pp. 56-62, doi: 10.1145/2844110.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015), "Algorithm aversion: people erroneously avoid algorithms after seeing them err", *Journal of Experimental Psychology: General*, Vol. 144 No. 1, pp. 114-126, doi: 10.1037/xge0000033.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2018), "Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them", *Management Science*, Vol. 64 No. 3, pp. 1155-1170, doi: 10.1287/mnsc.2016.2643.

European Parliament and Council of the European Union (2024), "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)", *Official Journal of the European Union*, OJ L, 2024/1689, 12 July, pp. 1-144, available at:https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Fanni, R., Steinkogler, V.E., Zampedri, G. and Pierson, J. (2023), "Enhancing human agency through redress in artificial intelligence systems", *AI and Society*, Vol. 38 No. 2, pp. 537-547, doi: 10.1007/s00146-022-01454-7.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50, doi: 10.1177/002224378101800104.

Greenberg, J. (1987), "A taxonomy of organizational justice theories", *Academy of Management Review*, Vol. 12 No. 1, pp. 9-22, doi: 10.2307/257990.

Hair, J.F., Risher, J.J., Sarstedt, M. and Ringle, C.M. (2019), "When to use and how to report the results of PLS-SEM", *European Business Review*, Vol. 31 No. 1, pp. 2-24, doi: 10.1108/ebr-11-2018-0203.

Haugeland, I.K.F., Folstad, A., Taylor, C. and Bjorkli, C.A. (2022), "Understanding the user experience of customer service chatbots: an experimental study of chatbot interaction design", *International Journal of Human-Computer Studies*, Vol. 161, 102788, doi: 10.1016/j.ijhcs.2022.102788.

Hellwig, P., Buchholz, V., Kopp, S. and Maier, G.W. (2023), "Let the user have a say: voice in automated decision-making", *Computers in Human Behavior*, Vol. 138, 107446, doi: 10.1016/j.chb.2022.107446.

Hellwig, P., Schmid-Palzer, J. and Maier, G.W. (2025), "Control! why employees should have a voice in automated decision-making", *Social Sciences and Humanities Open*, Vol. 12, 102113, doi: 10.1016/j.ssaho.2025.102113.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135, doi: 10.1007/s11747-014-0403-8.

Ho, C.C. and MacDorman, K.F. (2010), "Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices", *Computers in Human Behavior*, Vol. 26 No. 6, pp. 1508-1518, doi: 10.1016/j.chb.2010.05.015.

Hoff, K.A. and Bashir, M. (2015), "Trust in automation: integrating empirical evidence on factors that influence trust", *Human Factors*, Vol. 57 No. 3, pp. 407-434, doi: 10.1177/0018720814547570.

Hosseini Tabaghdehi, S.A. and Ayaz, O. (2025), "AI ethics in action: a circular model for transparency, accountability and inclusivity", *Journal of Managerial Psychology.* doi: 10.1108/jmp-03-2024-0177.

International Organization for Standardization (2023), *ISO/IEC 42001:2023 Information Technology—Artificial Intelligence—Management System*, ISO, Geneva.

Jabagi, N., Croteau, A.-M., Audebrand, L.K. and Marsan, J. (2025), "Do algorithms play fair? Analysing the perceived fairness of HR-decisions made by algorithms and their impacts on gig-workers", *International Journal of Human Resource Management*, Vol. 36 No. 2, pp. 235-274, doi: 10.1080/09585192.2024.2441448.

Klingbeil, A., Grützner, C. and Schreck, P. (2024), "Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI", *Computers in Human Behavior*, Vol. 160, 108352, doi: 10.1016/j.chb.2024.108352.

Köchling, A. and 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*, Vol. 13 No. 3, pp. 795-848, doi: 10.1007/s40685-020-00134-w.

Köchling, A., Wehner, M.C. and Rühle, S.A. (2025), "This (AI)n't fair? Employee reactions to artificial intelligence (AI) in career development systems", *Review of Managerial Science*, Vol. 19 No. 4, pp. 1195-1228, doi: 10.1007/s11846-024-00789-3.

Koh, L.Y., Cheh, Y.G., Yuen, K.F. and Wang, X. (2025), "Satisfaction with AI chatbots in crowdsourcing services: anthropomorphic and fairness perspectives", *Behaviour and Information Technology*, Vol. 44 No. 17, pp. 4198-4219, doi: 10.1080/0144929x.2025.2469660.

Kowald, D., Scher, S., Pammer-Schindler, V., Müllner, P., Waxnegger, K., Demelius, L., Fessl, A., Toller, M., Mendoza Estrada, I.G., Šimić, I., Sabol, V., Trügler, A., Veas, E., Kern, R., Nad, T. and Kopeinik, S. (2024), "Establishing and evaluating trustworthy AI: overview and research challenges", *Frontiers in Big Data*, Vol. 7, 1467222, doi: 10.3389/fdata.2024.1467222.

Langer, M. and 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*, Vol. 123, 106878, doi: 10.1016/j.chb.2021.106878.

Langer, M., Baum, K. and Schlicker, N. (2024), "Effective human oversight of AI-based systems: a signal detection perspective on the detection of inaccurate and unfair outputs", *Minds and Machines*, Vol. 35 No. 1, doi: 10.1007/s11023-024-09701-0.

Lee, M.K. (2018), "Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management", *Big Data and Society*, Vol. 5 No. 1, doi: 10.1177/2053951718756684.

Lee, J.D. and See, K.A. (2004), "Trust in automation: designing for appropriate reliance", *Human Factors*, Vol. 46 No. 1, pp. 50-80, doi: 10.1518/hfes.46.1.50.30392.

Leventhal, G.S., Karuza, J. and Fry, W.R. (1980), "Beyond fairness: a theory of allocation preferences", in Mikula, G. (Ed.), *Justice and Social Interaction*, Springer-Verlag, New York, NY, pp. 167-218.

Lind, E.A. (2001), "Fairness heuristic theory: justice judgments as pivotal cognitions in organizational relations", in Greenberg, J. and Cropanzano, R. (Eds), *Advances in Organizational Justice*, Stanford University Press, Stanford, CA, pp. 56-88.

Lind, E.A., Kanfer, R. and Earley, P.C. (1990), "Voice, control, and procedural justice: instrumental and noninstrumental concerns in fairness judgments", *Journal of Personality and Social Psychology*, Vol. 59 No. 5, pp. 952-959, doi: 10.1037/0022-3514.59.5.952.

Lippert, I., Alizadeh, A., Tarafdar, M., Mohlmann, M., Benlian, A., Parent-Rocheleau, X. and Stein, M.K. (2026), "One decade of algorithmic management research", *Business and Information Systems Engineering*, Vol. 68 No. 4, pp. 959-969, doi: 10.1007/s12599-026-00995-1.

Logg, J.M., Minson, J.A. and Moore, D.A. (2019), "Algorithm appreciation: people prefer algorithmic to human judgment", *Organizational Behavior and Human Decision Processes*, Vol. 151, pp. 90-103, doi: 10.1016/j.obhdp.2018.12.005.

Lu, L., McDonald, C., Kelleher, T., Lee, S., Chung, Y.J., Mueller, S., Vielledent, M. and Yue, C.A. (2022), "Measuring consumer-perceived humanness of online organizational agents", *Computers in Human Behavior*, Vol. 128, 107092, doi: 10.1016/j.chb.2021.107092.

Lucas, G.M., Becerik-Gerber, B. and Roll, S.C. (2024), "Calibrating workers' trust in intelligent automated systems", *Patterns*, Vol. 5 No. 9, 101045, doi: 10.1016/j.patter.2024.101045.

Majrashi, K. (2025), "Employees' perceptions of the fairness of AI-based performance prediction features", *Cogent Business and Management*, Vol. 12 No. 1, 2456111, doi: 10.1080/23311975.2025.2456111.

Malin, C.D., Fleiß, J., Seeber, I., Kubicek, B., Kupfer, C. and Thalmann, S. (2024), "The application of AI in digital HRM - an experiment on human decision-making in personnel selection", *Business Process Management Journal*, Vol. 30 No. 8, pp. 284-312, doi: 10.1108/bpmj-11-2023-0884.

Men, L.R., Zhou, A. and Tsai, W.H.S. (2022), "Harnessing the power of chatbot social conversation for organizational listening: the impact on perceived transparency and organization-public relationships", *Journal of Public Relations Research*, Vol. 34 Nos 1-2, pp. 20-44, doi: 10.1080/1062726x.2022.2068553.

Mirbabaie, M., Langer, M., Rieskamp, J. and Hofeditz, L. (2026), "Transparency fallacy: perceived fairness in algorithmic management", *Business and Information Systems Engineering*, Vol. 68 No. 4, pp. 829-850, doi: 10.1007/s12599-025-00963-1.

Mori, M., MacDorman, K.F. and Kageki, N. (2012), "The uncanny valley", *IEEE Robotics and Automation Magazine*, Vol. 19 No. 2, pp. 98-100, doi: 10.1109/MRA.2012.2192811.

Moritz, J.M. and Wehner, M.C. (2026), "Justice evaluations of algorithmic management: the role of prior discrimination experience", *Business and Information Systems Engineering*, pp. 1-23, doi: 10.1007/s12599-026-01002-3.

Narayanan, D., Nagpal, M., McGuire, J., Schweitzer, S. and De Cremer, D. (2024), "Fairness perceptions of artificial intelligence: a review and path forward", *International Journal of Human-Computer Interaction*, Vol. 40 No. 1, pp. 4-23, doi: 10.1080/10447318.2023.2210890.

National AI Centre (2026), *Guidance for AI Adoption: Implementation Guidance*, Australian Government, Canberra, 5 May.

Newman, D.T., Fast, N.J. and Harmon, D.J. (2020), "When eliminating bias is not fair: algorithmic reductionism and procedural justice in human resource decisions", *Organizational Behavior and Human Decision Processes*, Vol. 160, pp. 149-167, doi: 10.1016/j.obhdp.2020.03.008.

Okwir, S., Nudurupati, S.S., Ginieis, M. and Angelis, J. (2018), "Performance measurement and management systems: a perspective from complexity theory", *International Journal of Management Reviews*, Vol. 20 No. 3, pp. 731-754, doi: 10.1111/ijmr.12184.

Pan, Y., Froese, F.J. and Xue, S. (2026), "The role of AI in performance appraisal: a mixed-method study of employee experience through a relational lens", *Human Resource Management*, Vol. 65 No. 3, pp. 781-798, doi: 10.1002/hrm.70049.

Pelau, C., Dabija, D.C. and Ene, I. (2021), "What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry", *Computers in Human Behavior*, Vol. 122, 106855, doi: 10.1016/j.chb.2021.106855.

Porra, J., Lacity, M. and Parks, M.S. (2020), "Can computer-based human-likeness endanger humanness? A philosophical and ethical perspective on digital assistants expressing feelings they cannot have", *Information Systems Frontiers*, Vol. 22 No. 3, pp. 533-547, doi: 10.1007/s10796-019-09969-z.

Qin, S., Jia, N., Luo, X., Liao, C. and Huang, Z. (2023), "Perceived fairness of human managers compared with artificial intelligence in employee performance evaluation", *Journal of Management Information Systems*, Vol. 40 No. 4, pp. 1039-1070, doi: 10.1080/07421222.2023.2267316.

Revillod, G. (2025), "Trust influence on AI HR tools perceived usefulness in Swiss HRM: the mediating roles of perceived fairness and privacy concerns", *AI and Society*, Vol. 40 No. 6, pp. 4789-4822, doi: 10.1007/s00146-025-02216-x.

Rhim, J., Kwak, M., Gong, Y. and Gweon, G. (2022), "Application of humanization to survey chatbots: change in chatbot perception, interaction experience, and survey data quality", *Computers in Human Behavior*, Vol. 126, 107034, doi: 10.1016/j.chb.2021.107034.

Sarstedt, M., Ringle, C.M., Cheah, J.-H., Ting, H., Moisescu, O.I. and Radomir, L. (2020), "Structural model robustness checks in PLS-SEM", *Tourism Economics*, Vol. 26 No. 4, pp. 531-554, doi: 10.1177/1354816618823921.

Schuetzler, R.M., Grimes, G.M. and Giboney, J.S. (2020), "The impact of chatbot conversational skill on engagement and perceived humanness", *Journal of Management Information Systems*, Vol. 37 No. 3, pp. 875-900, doi: 10.1080/07421222.2020.1790204.

Shin, D. (2020), "User perceptions of algorithmic decisions in the personalized AI system: perceptual evaluation of fairness, accountability, transparency, and explainability", *Journal of Broadcasting and Electronic Media*, Vol. 64 No. 4, pp. 541-565, doi: 10.1080/08838151.2020.1843357.

Shin, D. (2021), "The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI", *International Journal of Human-Computer Studies*, Vol. 146, 102551, doi: 10.1016/j.ijhcs.2020.102551.

Shin, D. (2022), "The perception of humanness in conversational journalism: an algorithmic information-processing perspective", *New Media and Society*, Vol. 24 No. 12, pp. 2680-2704, doi: 10.1177/1461444821993801.

Shin, D. and Park, Y.J. (2019), "Role of fairness, accountability, and transparency in algorithmic affordance", *Computers in Human Behavior*, Vol. 98, pp. 277-284, doi: 10.1016/j.chb.2019.04.019.

Shin, H.H., Choi, S. and Kim, H. (2025), "Artificial Intelligence (AI) in human resource management (HRM): a driver of organizational dehumanization and negative employee reactions", *International Journal of Hospitality Management*, Vol. 131, 104230, doi: 10.1016/j.ijhm.2025.104230.

Shulner-Tal, A., Kuflik, T., Kliger, D. and Mancini, A. (2025), "Who made that decision and why? Users' perceptions of human versus AI decision-making and the power of explainable-AI", *International Journal of Human-Computer Interaction*, Vol. 41 No. 7, pp. 4230-4247, doi: 10.1080/10447318.2024.2348843.

Spatola, N. (2024), "The efficiency-accountability tradeoff in AI integration: effects on human performance and over-reliance", *Computers in Human Behavior: Artificial Humans*, Vol. 2 No. 2, 100099, doi: 10.1016/j.chbah.2024.100099.

Sullivan, Y., de Bourmont, M. and Dunaway, M. (2022), "Appraisals of harms and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems", *Annals of Operations Research*, Vol. 308 No. 1, pp. 525-548, doi: 10.1007/s10479-020-03702-9.

Tarafdar, M., Page, X. and Marabelli, M. (2023), "Algorithms as co-workers: human-algorithm role interactions in algorithmic work", *Information Systems Journal*, Vol. 33 No. 2, pp. 232-267, doi: 10.1111/isj.12389.

Thibaut, J. and Walker, L. (1975), *Procedural Justice: a Psychological Analysis*, Erlbaum, Hillsdale, NJ.

Tinguely, P.N., Lee, J. and He, V.F. (2023), "Designing human resource management systems in the age of AI", *Journal of Organization Design*, Vol. 12 No. 4, pp. 263-269, doi: 10.1007/s41469-023-00153-x.

Vantrepotte, Q., Berberian, B., Pagliari, M. and Chambon, V. (2022), "Leveraging human agency to improve confidence and acceptability in human-machine interactions", *Cognition*, Vol. 222, 105020, doi: 10.1016/j.cognition.2022.105020.

Varma, A., Pereira, V. and Patel, P. (2024), "Artificial intelligence and performance management", *Organizational Dynamics*, Vol. 53 No. 1, 101037, doi: 10.1016/j.orgdyn.2024.101037.

Wang, N., Zhang, X., Li, S. and Gao, X. (2025), "Applications of artificial intelligence in enterprise human resource management", *Information Resources Management Journal*, Vol. 38 No. 1, pp. 1-19, doi: 10.4018/IRMJ.389707.

Wesche, J.S., Hennig, F., Kollhed, C.S., Quade, J., Kluge, S. and Sonderegger, A. (2024), "People's reactions to decisions by human versus algorithmic decision-makers: the role of explanations and type of selection tests", *European Journal of Work and Organizational Psychology*, Vol. 33 No. 2, pp. 146-157, doi: 10.1080/1359432x.2022.2132940.

Wood, R.E. (1986), "Task complexity: definition of the construct", *Organizational Behavior and Human Decision Processes*, Vol. 37 No. 1, pp. 60-82, doi: 10.1016/0749-5978(86)90044-0.

Xu, Y., Shieh, C.H., van Esch, P. and Ling, I.L. (2020), "AI customer service: task complexity, problem-solving ability, and usage intention", *Australasian Marketing Journal*, Vol. 28 No. 4, pp. 189-199, doi: 10.1016/j.ausmj.2020.03.005.

Yu, J., Ma, Z. and Zhu, L. (2025), "The configurational effects of artificial intelligence-based hiring decisions on applicants' justice perception and organisational commitment", *Information Technology and People*, Vol. 38 No. 2, pp. 553-581, doi: 10.1108/itp-04-2022-0271.

Zdravković, M. and Panetto, H. (2022), "Artificial intelligence-enabled enterprise information systems", *Enterprise Information Systems*, Vol. 16 No. 5, 1973570, doi: 10.1080/17517575.2021.1973570.

Corresponding author
Md Irfanuzzaman Khan can be contacted at: irfan.khan@canberra.edu.au